Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Air Force Personnel and Training Research Center

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *AF Personnel and Training Research Center*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
NEAL E. MILLER, *Yale University*
KENNETH W. SPENCE, *State University of Iowa*

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Assistant Managing Editor
Editorial Staff: SADIE J. DOYLE, FRANCES H. CLARK, BARBARA CUMMINGS

CONTENTS

Training Under Two Drives, Alternately Present, vs. Training Under a Single Drive: L. W. PORTER AND N. E. MILLER 1

The Influence of Amount of Prerest Learning on Reminiscence Effects in Paired-Associate Learning: D. A. RILEY 8

Effect of Intertrial Interval on Acquisition and Extinction of a Running Response: J. W. COTTON AND D. J. LEWIS 15

The Electrical Phosphene Threshold as a Measure of Retinal Induction and Visual Organization: R. M. MICHAELS 21

Learning the Statistical Properties of the Input in Pursuit Tracking: E. C. POULTON 28

Studies of Distributed Practice: XV. Verbal Concept Learning as a Function of Intralist Interference: B. J. UNDERWOOD 33

Differential Sensitivity in Gustation: H. G. SCHUTZ AND F. J. PILGRIM 41

Effect of Reduced Exposure Duration on Brightness Constancy: H. LEIBOWITZ AND P. CHINETTI 49

The Relationship of Performance Level to Level of Arousal: R. G. STENNETT 54

Learning Through Stimulus Satiation: M. R. DENNY 62

Factors in the Recovery From Approach-Avoidance Conflict: M. M. BERKUN 65

Meaning Established by Classical Conditioning: C. K. STAATS AND A. W. STAATS 74

American Psychological Association

Vol. 54 No. 1 July 1957

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal, experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research effort is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, 53, 1–2).

Address all articles submitted for publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, P. O. Box 1557, Lackland Air Force Base, Texas. All manuscripts should be submitted in two complete copies; only the original copy will be returned if the article is not accepted for publication. Figures may be original line drawings or glossy photographic prints, but should not exceed 8½ by 11 in. Care should be taken to insure legibility of labels and critical features of drawings when reduced to a single-column or (exceptionally) a double-column figure in the JOURNAL.

To save printing costs supplementary material to articles in this JOURNAL are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1957 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 54, NO. 1 JULY, 1957

TRAINING UNDER TWO DRIVES, ALTERNATELY PRESENT, VS. TRAINING UNDER A SINGLE DRIVE [1]

LYMAN W. PORTER [2] AND NEAL E. MILLER

Yale University

To date, most experiments on learning have used but a single relevant drive. A few experiments have studied the effects of two drives, one of which was an irrelevant drive because its goal object was not present (e.g. **1, 2, 3, 6, 7**). Very few studies (e.g. **4, 5, 8**) have used two relevant drives. Those that have used more than one relevant drive have combined them simultaneously but never successively. However, combining drives simultaneously involves the possibility of physiological interaction between the drives themselves (e.g., dehydration interfering with hunger), a complicating factor which is avoided if the two drives are used successively.

In many life situations a given habit is motivated by more than one drive. For example, the habit of working for money has been learned in response to a number of drives which were present either simultaneously or successively. In order to better understand motivation in many social and clinical situations we need to know more about how different drives combine. Are extra trials with a different drive somewhat less or somewhat more effective than the same number of extra trials with the same drive, or do they produce interference which weakens instead of strengthens the habit? How do the results depend on the level of learning, the particular drives which are combined, and the conditions under which the habit is tested?

A theoretical analysis shows that training under two drives, alternately present, might be expected to increase performance during subsequent satiation of both drives. Animals trained under the two drives, alternately present, will be reinforced for performing during the absence of cues from each of the drives. Animals trained under a single drive will never have been reinforced for performing in the absence of cues from that drive. On this basis it might be expected that training under two drives would be more efficient than training under a single drive.

[1] This investigation was supported by a research grant, M647, from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service. The main findings were presented at the 1955 Meeting of the Eastern Psychological Association.

[2] Now at the University of California.

On the other hand, it is possible that the anticipatory goal responses of the two drives interfere with each other somewhat, or that a certain amount of habit strength is lost in the generalization from one drive stimulus to the other. On this basis, it might be expected that training under two drives would be less efficient than training under one drive.

Without a quantitative theory one cannot predict which of these two opposing effects will be stronger, but if it is assumed that the drive stimulus during training is a significant fraction of the cue pattern and that the cues from the different drives are distinctive, the training with two drives (instead of only one) should be more beneficial (or less detrimental) when extinction tests are given during satiation than when they are given with a relevant drive.

The purpose of the following experiments was to compare the effects of training under two drives (hunger and thirst) alternately present with those of training under one drive, either hunger or thirst. Two measures were used: (*a*) *satiated extinction* (both drive and reward absent), and (*b*) *regular extinction* (drive present but reward absent).

Experiment I

Method

Subjects and apparatus.—The *Ss* were 64 male albino rats of the Sprague-Dawley strain, 70–90 days old at the start of the experiment. The apparatus consisted of six modified Skinner boxes. A ¼-in.-thick brass stirrup bar extended ½ in. into the end of the box, and beneath it a 1½-in.-square section of the floor was removed. An automatic feeding device presented an inverted soda bottle cap that contained either food pellets or drops of water, ½ in. below this opening in the floor. A pressure of 15 gm. on the bar activated a microswitch which: (*a*) registered the response on an electric counter, and (*b*) operated the feeding device to bring a new loaded cup into place. During extinction, responses were counted but the feeding device did not operate.

Groups and procedure.—In order to equate the strength of motivation in the single- and double-drive groups during training, all *Ss* were kept on the same hunger and thirst regimen throughout the 24 pretest days. All of these *Ss* were made hungry and thirsty on alternate days. The drive-training differences among the groups were produced by training them in the apparatus during certain of these days according to the procedures described below.

Double drive, 12-day group (DD-12): The 16 *Ss* in this group received training in the apparatus alternately under hunger and thirst. That is, half of this group (eight *Ss*) were put in the apparatus hungry one day, then thirsty the next, and then had two intervening days of no training before the cycle was repeated. The other half of the group received the reverse order of the cycle (i.e., thirst-hunger) as a control for any sequence effects. For all 16 *Ss* in Group DD-12 the 24 pretest days involved six days training with hunger and six days with thirst.

Single drive, 12-day group (SD-12): The 16 *Ss* in this group received training in the apparatus under one drive, either hunger or thirst. Eight *Ss* were run only when hungry and the eight others only when thirsty. The *Ss* in Group SD-12 were run every other day during the 24 pretest days, and were always run under the same drive. Thus, these *Ss* received as much training in the apparatus (12 days) as Group DD-12, but had twice the training on one of Group DD-12's two drives.

Single drive, 6-day group (SD-6): The 16 *Ss* in this group differed from Group SD-12 in that they were run every other day during the final 12 pretesting days, and so received only six days of training in the apparatus (instead of 12), or as much training on its one drive as Group DD-12 received on *one* of its two drives.

Distribution control groups.—In order to measure whether the difference in distribution of training days between Group DD-12 and the two SD groups might be causing any differences in the groups during testing, a control group of 16 *Ss* was run. Eight of these *Ss* received training in the apparatus according to the schedule for Group DD-12 (i.e., alternating two training days with two nontraining days), and the other eight received training according to the schedule for the SD groups (i.e., alternating one training day with one nontraining day). Each of these subgroups was divided into two subsubgroups, with four *Ss* always hungry when trained and four always thirsty when trained.

Control of drive.—The method for producing hunger for the three experimental groups involved removing the food that remained from

TABLE 1

NUMBER OF BAR PRESSES DURING 30-MIN. SATIATED EXTINCTION TEST IN EXP. I

Group	N	Number of Bar Presses					
		Hunger Trained		Thirst Trained		Total N	
		Mean	SD	Mean	SD	Mean	SD
DD-12	16					47.0	18.7
SD-12	8 + 8	38.8	16.7	43.6	14.6	41.2	15.4
SD-6	8 + 8	40.0	20.9	28.1	13.0	34.1	17.9
Dist. DD	8					38.6	27.1
Dist. DD	8					40.8	15.7

the previous 23½ hr. of ad lib. feeding while *S* was being run in the apparatus (under thirst), and depriving *S* of food for the next 24 hr. when only ad lib. water was available. Since it was considered desirable to try to equate the strength of the hunger and thirst drives as nearly as possible, a slightly different procedure was adopted for controlling the thirst drive. To make *Ss* thirsty, food and water were given together for 30 min. following running in the apparatus (while hungry), then water was removed and only food remained for the next 23½ hr. These two slightly different procedures were found to produce approximately equal bar-pressing rates in a prior exploratory study.

The distribution control group received a different type of feeding and drinking schedule. Half of this group was always hungry and half always thirsty. Hungry *Ss* were given 30 min. of free access to food following running, but no water was available during the 30 min. During the remainder of the 23½ hr. before the next feeding only water was available. The thirsty *Ss* were given 30 min. of food and water together following training in the apparatus (or at that time of day on nontraining days), and then only food was available for the next 23½ hr.

Training.—For all *Ss* the first training day was used to teach the bar-pressing response in the apparatus. This was accomplished by having *E* "shape" the response until *S* was able to press the bar 15 times unaided by *E*. On all remaining training days *Ss* were allowed to bar press for 15 reinforcements delivered on a 100% reinforcement schedule. Thus, Groups DD-12 and SD-12 received a total of 165 reinforced bar presses following the initial training day, and Group SD-6 received a total of 75. The amount of one reinforcement on hunger days was two pellets, weighing 46 mg. each, and on thirst days was three drops of water (.15 cc.) from an eye-dropper.

Testing.—The purpose of this first experiment was to test *Ss* under satiated extinction. Therefore, all *Ss* in all groups were satiated by giving them ad lib. dry food and water for 23½ hr. following their final training day. Then 30 min. prior to being put in the apparatus for the extinction test the dry food was replaced by wet mash and the water bottle was shaken to draw *S*'s attention. In the test situation the number of bar presses made in each 3-min. interval was recorded for 30 min. with the delivery device inoperative.

Results

The mean number of bar presses made in the 30-min. testing period are presented in Table 1. The two distribution control groups were not significantly different from each other. When the SD groups are split into their hunger-trained and thirst-trained subgroups, it is found that for Group SD-12 hunger and thirst were nearly equivalent in their training effects on the satiated extinction. In Group SD-6 the hunger subgroup made more bar presses than the thirst subgroup although the difference was not reliable.

While the *Ss* given additional days of training on the single drive (SD-12) did perform better on the satiated extinction test than those with fewer such trials (SD-6), the difference was not reliable ($P = .24$).[3] The addi-

[3] All *P* values in this paper are for two-tailed tests; in some instances it might be considered permissible to divide these values by two since the differences are in the predicted direction.

TABLE 2

NUMBER OF BAR PRESSES DURING 30-MIN. REGULAR EXTINCTION TEST IN EXP. II

Group	*N*	Number of Bar Presses					
		Hunger Tested		Thirst Tested		Total *N*	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
DD-12	6 + 6	81.5	10.5	66.3	19.8	73.9	17.1
SD-12	4 + 5	59.0	15.9	60.6	15.9	59.9	14.9

tional days of training under a second drive (DD-12 vs. SD-6) definitely did not produce any interference, but instead, increased performance on the satiated extinction test ($P = .05$). When the total number of trials was equated (DD-12 vs. SD-12), the double-drive group performed somewhat better, but not reliably so ($P = .35$).

EXPERIMENT II

Method

The purpose of this experiment was to train *Ss* under exactly the same procedures as used in Exp. I, but to test them under regular drive extinction. Therefore, everything was kept the same in this experiment except: (*a*) 32 *Ss* were used, 16 in a Group DD-12 and 16 in a Group SD-12 that corresponded to the DD-12 and SD-12 groups of the previous experiment. No SD-6 or Distribution Control groups were run. (*b*) The *Ss* in both groups were extinguished on the same strength of drive as that used during training. Half of Group DD-12 was extinguished under thirst, and the other half under hunger. In Group SD-12, *Ss* were extinguished under their particular training drive. Other testing conditions were the same as for Exp. I.

Results

The mean number of bar presses made in the 30-min. test period for *Ss* in Exp. II are presented in Table 2. (The *Ns* are reduced from 16 each in the SD-12 and DD-12 groups to 9 in the SD-12 group and 12 in the DD-12 group due to deaths of *Ss* during the training period.) When Group SD-12 is divided into thirst-trained-and-tested and hunger-trained-and-tested subgroups, it is seen that they performed virtually the same on the test, the means being 60.6 and 59.0. When Group DD-12 is divided into the six *Ss* extinguished on thirst vs. the six *Ss* extinguished on hunger (both groups having received equal hunger and thirst training) it is seen that *Ss* extinguished on hunger performed better, but not reliably so.

In this experiment using the regular extinction test, Group DD-12 again performed slightly but not reliably better than Group SD-12 ($P = .07$).

EXPERIMENT III

Method

The purposes of Exp. III were: (*a*) to train *Ss* under procedures nearly identical with the first two experiments, but to use both satiated and regular extinction tests on *Ss* randomly selected from the same population; and (*b*) to retrain the *Ss* following their test period and then retest them again under the other condition. Everything was kept the same in this experiment except: (*a*) one pellet of food (46 mg.) was used as reward for each bar press on hunger days instead of two pellets; (*b*) in the extinction tests the apparatus *did* deliver the cups following each bar press, but there was no food or water in the cups; and (*c*) 64 *Ss* were used, 32 in a DD-12 group and 32 in a SD-12 group. Two testing periods—satiated extinction and regular extinction—were given in the experiment. Half of Group DD-12 (eight hunger-thirst *Ss* and eight thirst-hunger *Ss*) and half of Group SD-12 (eight hunger *Ss* and eight thirst *Ss*) received the satiated extinction test first and the regular extinction test second. The other halves of both groups received the reverse

TABLE 3

NUMBER OF BAR PRESSES DURING 30-MIN. TESTS IN EXP. III

Group	*N*	Satiation Test				Regular Extinction Test			
		First		Second		First		Second	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
DD-12	16	99.1	37.1	51.9	28.5	162.2†	57.6	110.5	56.9
SD-12	16	74.4*	28.6	45.2	24.4	168.2‡	72.0	92.3	32.2

* Mean for 8 hunger-trained *Ss* = 91.1 (*SD* = 25.9); for 8 thirst-trained *Ss* = 57.8 (*SD* = 20.9).
† Mean for 8 hunger-tested *Ss* = 165.0 (*SD* = 60.6); for 8 thirst-tested *Ss* = 159.5 (*SD* = 58.6).
‡ Mean for 8 hunger-trained-and-tested *Ss* = 198.5 (*SD* = 76.0); for 8 thirst-trained-and-tested *Ss* = 137.9 (*SD* = 56.8).

order of testing. Between the two testing periods the feeding regimen was re-established and all *Ss* received two retraining days in the apparatus just as if no testing period had taken place. The second testing period took place 48 hr. after the final of the two retraining days.

Results

The mean number of bar presses made in the 30-min. test periods for *Ss* in Exp. III are presented in Table 3. When the SD-12 *Ss* are divided into hunger-trained and thirst-trained *Ss* the hunger-trained-and-tested group is superior on both regular extinction ($P = .05$) and satiated extinction ($P = .01$). When Group DD-12 is divided into hunger-tested and thirst-tested *Ss* on regular extinction, there is no significant difference between these subgroups.

In the combined first and second satiated extinction tests Group DD-12 made significantly more bar presses ($P = .03$) than Group SD-12, the means being 99.1 and 74.4, respectively, for those *Ss* whose first test was satiation and 51.9 and 45.2 for those *Ss* whose second test was satiation. On the first satiation test, however, Group DD-12 was not significantly superior to the better of the SD-12 subgroups, the hunger-trained SD-12 *Ss*.

On the regular extinction tests, there was no clear superiority of either group. For *Ss* given the regular extinction test first, the combined Group SD-12 was slightly superior to Group DD-12, but for *Ss* given the regular extinction test second, the difference was reversed. Neither of these differences was statistically reliable.

DISCUSSION

From the results of Exp. I and II it appeared that the feeding regimen was such as to ensure about equal performance by hunger-trained and thirst-trained SD *Ss* on both regular and satiated extinction. However, for some unaccountable reason, these different SD subgroups were not equal in Exp. III. The hunger-trained *Ss* were reliably superior on both regular and satiated extinction. Although this fact complicates the interpretations of Exp. III, Group DD-12 can be compared directly with the combined Group SD-12 because half of the *Ss* in the former group were trained under hunger and half under thirst.

The results of the regular extinction tests (Exp. II and part of Exp. III) were inconclusive. In both experiments there was a statistically unreliable tendency for Group DD-12 to show greater resistance to extinction compared to the combined Group SD-12. In Exp. III, Group DD-12 was inferior to the better

of the SD-12 subgroups, the SD-hunger group. These results indicate that when testing was carried out under one of the drives used during training, any advantage the DD-12 group gained by being trained under more than one condition (i.e. under a hunger-drive condition and under a thirst-drive condition) was balanced by the fact that the SD-12 groups had twice as much training on the particular drive present in testing.

The satiated extinction results showed that additional trials with a different drive reliably increased resistance to extinction and probably were slightly more effective than additional trials with the same drive. The latter difference (DD-12 superior to SD-12) came out in the same direction in both experiments, being unreliable in Exp. I and reliable in Exp. III. However, in Exp. III the DD-12 group was not reliably superior to the better of the two SD-12 subgroups, the hunger one. This latter result is similar to that found by Powlowski (5), who combined two drives simultaneously and found that the combination was not superior to the stronger of the two drives used singly.

The prediction from the theoretical analysis presented in the introduction was weakly confirmed on the first extinction tests in Exp. III, in that Group DD-12 showed relatively more superiority to the combined Group SD-12 under the condition of satiation than it did under the condition of regular drive extinction, although this interaction between drive and test conditions was not statistically reliable ($P = .24$).

It is probable that the *S*s had not reached their asymptote of training under either drive condition (especially Group DD-12, which had one-half as much training on each drive as the SD-12 groups). If training had been carried out to asymptote it is possible that the superiority of Group DD-12 would have been greater. In any event, the results of these experiments show that a group trained alternately under two drives is no worse in either test than a group given the same over-all amount of training but with all training on one drive. If testing is carried out under satiated extinction conditions, the group trained under two drives appears to be as good as the superior SD-12 subgroup and better than the inferior SD-12 subgroup.

Summary

The purpose of these experiments was to compare the effects of training under two drives (hunger and thirst), alternately present, with those of training under one drive (either hunger or thirst). Performance was measured under two conditions: (*a*) satiated extinction (both drive and reward absent), and (*b*) regular extinction (drive present but reward absent).

Three experiments were performed, using a total of 149 male albino rats. Training and testing were carried out in modified Skinner boxes, and the response measure used in extinction testing was the number of bar presses made during a 30-min. period. The main results were as follows:

1. Additional training trials with a different drive did not produce any interference, but instead improved performance.

2. When the total number of trials was held constant, rats with these trials given alternately on hunger and thirst were as good as those given all of their training on either hunger or thirst in their resistance to regular extinction, and tended to be somewhat superior in their resistance to satiated extinction.

References

1. Kendler, H. H. Drive interaction. I. Learning as a function of the simultaneous presence of hunger and thirst drives. *J. exp. Psychol.*, 1945, **35**, 96–109.
2. Kendler, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, **36**, 212–220.
3. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
4. Muenzinger, K. F., & Fletcher, F. M. Motivation in learning. VI. Escape from electric shock compared with hunger-food tension in the visual dis-

crimination habit. *J. comp. Psychol.*, 1936, **22**, 79–91.

5. Powloski, R. F. The effects of combining hunger and thirst motives in a discrimination habit. *J. comp. physiol. Psychol.*, 1953, **46**, 434–437.
6. Siegel, P. S. Alien drive, habit strength, and resistance to extinction. *J. comp. Psychol.*, 1946, **39**, 307–317.
7. Webb, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, **39**, 1–14.
8. Wike, E. L., & McNamara, H. J. A quest for the generalized conditioned reinforcer. *Psychol. Rep.*, 1955, **1**, 83–91.

(Received July 5, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

THE INFLUENCE OF AMOUNT OF PREREST LEARNING ON REMINISCENCE EFFECTS IN PAIRED-ASSOCIATE LEARNING

DONALD A. RILEY

University of California

In two previous papers (**7, 8**) evidence was presented that in the learning of paired-associate nonsense syllables, a single 2-min. rest filled by color naming can improve postrest performance.

In the previous experiments *Ss* learned two sets of paired-associate nonsense syllables, the first (List A), for four trials, the second (List B), for six trials prior to the rest. The rest facilitated the further performance of either list when compared to a nonrest control. However, when prerest training was on but a single list, the rest did not aid the subsequent performance. The results of these experiments did not appear compatible with the hypothesis that reminiscence results from a release from performance-inhibition, nor from differential forgetting, as supported by Easley (**1**) and by McGeoch and Irion (**5**).

The present experiment, a simple extension of the previous studies, investigates the effect of varying the number of prerest List B trials on List B postrest performance. The first list is held constant at four trials; the rest occurs after 0, 6, 11, or 16 trials of the second list in the four different conditions. In a final condition of the experiment the number of List A trials is increased to 10, and there are 11 prerest List B trials as in the second condition. This condition was run to check on the possibility that the rest aids further performance when prerest learning of the two lists is of approximately equal strength. As will be shown, the experiment as a whole supports this interpretation.

Method

Conditions and subjects.—The experimental conditions are designated E4-0, E4-6, E4-11, E4-16, and E10-11. In each case, the first number is the number of List A trials, the second is the number of List B trials prior to the rest. There were 20 *Ss* in each experimental condition. Each experimental *S* was matched with a control *S* who did not receive a rest interval. Matching of *Ss* was done on the basis of prerest performance. Two criteria were used: the number of items correct on all prerest trials and the number of items correct on the last prerest trial. Since all control *Ss* in the first four conditions were treated identically, only enough were run to insure that in each condition, each *S* had a closely matching control. That is, the same control *S* might be used in two different conditions if he were the best match for an experimental *S* in each condition. In order to insure close matching, 40 control *Ss* were run. The 20 control *Ss* that matched the *Ss* of E10-11 were run separately, and practiced List A for 10 trials before switching to List B. The control conditions are designated C4-0, C4-6, C4-11, C4-16, and C10-11. All *Ss* were students at the University of California.

Procedure.—The experimental procedure was essentially the same as that used earlier (**7**, Exp. 1). After instructions on paired-associate learning and warning that some items might change, *Ss* were instructed to learn whatever appeared. Pretraining for all *Ss* consisted of memorizing for 8 trials, by the anticipation method, 8 pairs of 3-letter words. Each stimulus word appeared for 2 sec. and was followed for 2 sec. by both the stimulus and response words. The pairs appeared in six different orders. An 8-sec. rest followed each trial except Trial 4, after which *Ss* named colors from a board for 2 min. Colors were named at a 2-sec. rate. The item presentation rates and color naming procedures used in pretraining were also used in the main experiment.

TABLE 1

MEAN ITEMS CORRECT ON PREREST AND POSTREST TRIALS ON LIST B

Cond.	Prerest Trials				Postrest Trials					
	All		Last		Trial 1		Trial 2		Trial 3	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
4-0										
E (Rest)	.53	.57	.80	.98	—	—	.15	.36	.30	.56
Control	.51	.54	.80	.87	—	—	.10	.30	.20	.40
Diff.	.02		.00		—		.05		.10	
4-6										
E (Rest)	.51	.52	1.05	1.07	1.50	1.28	2.10	1.26	2.35	1.59
Control	.53	.57	1.10	1.05	1.50	1.50	1.60	1.32	2.15	1.76
Diff.	−.02		−.05		.00		.50		.20	
4-11										
E (Rest)	1.56	.99	2.90	1.73	2.65	1.59	3.35	1.71	3.35	2.03
Control	1.54	1.05	3.00	1.76	3.50	1.86	3.40	1.85	3.45	1.86
Diff.	.02		−.10		−.85**		−.05		−.10	
4-16										
E (Rest)	2.12	1.25	3.80	1.89	3.60	2.26	4.40	1.92	4.75	2.13
Control	2.10	1.10	4.00	1.95	4.35	1.85	4.25	1.66	4.45	2.16
Diff.	.02		−.20		−.75*		.15		.30	
10-11										
E (Rest)	1.33	.80	2.50	1.67	2.50	1.95	3.70	1.91	4.05	1.98
Control	1.32	.86	2.70	1.49	2.65	1.55	2.65	1.69	3.15	2.24
Diff.	.01		−.20		−.15		1.05**		.90*	

Note.—Prerest values in Cond. 4-0 are for List A.
* $P = .05$.
** $P = .02$.

About 2 min. after pretraining was concluded, the main experiment began. Lists of paired-associate nonsense syllables previously described (**7**) were used. The stimulus items and the response items both averaged 15% association value (**2**) and were maximally dissimilar in accord with rules described in the previous paper. As in the earlier experiments, there were two 8-item sets of paired associates, the sets having stimulus items in common, but different response items. The *Ss* first memorized List A for 4 trials. Then, following the customary 8-sec. break, and with no specific warning, List B appeared. The *Ss* then memorized List B until one perfect recitation was achieved. Four different orders of List A and six different orders of List B were used. In all experimental conditions, every trial but one was followed by an 8-sec. pause. The one exception was the 2-min. color naming interval which followed the appropriate List B trial. In the control conditions all trials were followed by an 8-sec. interval.

The procedure in Cond. 10-11 differed from the above description in two respects. First, instead of 4 List A trials there were 10. Second, in order to insure equality of learning on Lists A and B prior to the rest, *Ss* were used only if the total number of List A syllables correct was within one of the number of prerest List B syllables correct. This criterion was met both in the experimental and control condition.

RESULTS

Prerest scores.—The prerest results are shown on the upper part of Table 1. The closeness of scores both on the means for all List B trials, and on the last prerest trial, merely reflects the fact that the correspondence was forced. The greatest difference that showed any regularity was the last prerest trial of Cond. 10-11. Here, four control *Ss* anticipated one

TABLE 2

POSTREST MEAN DIFFERENCES AND $\sigma_{diff.}$ AT SYLLABLE CRITERIA 3 THROUGH 8 IN COND. 4-0, 4-6, AND 10-11

Criterion	Cond. 4-0			Cond. 4-6			Cond. 10-11		
	Pairs	Mean Diff.	P	Pairs	Mean Diff.	P	Pairs	Mean Diff.	P
3	40	2.68±1.23	.05	17	3.29±1.72	.10	9	2.33±1.69	.20
4	40	4.08±1.69	.01	20	3.60±1.84	.05	15	6.80±2.87	.05
5	40	4.10±2.22	.10	20	4.10±2.40	.10	17	6.82±2.85	.05
6	40	5.28±2.64	.05	20	7.10±2.87	.05	17	6.53±3.66	.10
7	39	4.33±3.32	.20	20	9.20±3.34	.02	18	4.94±2.93	.20
8	32	1.56±3.91	.70	16	8.44±3.78	.05	15	5.40±2.83	.10

more syllable than the corresponding experimental *Ss*.

Postrest scores.—The effects of the rest interval in the different conditions are shown in Tables 1 and 2 and in Fig. 1. Table 1 shows for each condition, the experimental and control means on the first three postrest trials and the significance tests for the difference between these values.

Figure 1 allows direct comparison of all the conditions. It shows the postrest mean differences between the rest and nonrest conditions in number of trials taken to reach each successive syllable criterion. It will be seen that all experiments are not represented at all criterion levels. This is most notable in Cond. 4-16 in which most pairs of *Ss* had achieved the earlier criteria before the rest. Values are charted only where a sufficient number of *Ss* are represented to give some credence to these values. The small numbers by each point designate the numbers of pairs entering into the mean difference represented by the point. The variations in the number of pairs are the result of two factors. Some pairs of *Ss* learned too rapidly to be included in the earlier criteria, while a few pairs learned too slowly to complete all syllables by the end of the allotted period, and they are not represented in the later criteria. This second factor is quite insignificant except at 8 correct.

Table 2 presents the mean difference between the E and C groups at syllable Criteria 3 through 8 and the corresponding tests of significance for the three experiments showing a beneficial effect of the rest. Since these values represent mean differences in total trials to reach each criterion, each value is partially determined by the preceding value. As a consequence of this, the *t* tests may capitalize on chance; the significance values must be treated with caution. Results from this table are discussed below. Each of the conditions will now be considered in detail.

Condition 4-0.—Cond. 4-0 was originally completed with 20 pairs of *Ss*, and showed a beneficial effect of the rest which was significant at the .10 level of confidence at a criterion of 5 correct. Because of the inconclusiveness of the experiment it was repeated, with similar results. The data shown are for the two experiments combined. The closeness of the results in the original experiment and in the repetition, and the fact that the combined values show a statistically significant effect at 4 correct, support the conclusion that there is a slight beneficial effect of the rest when it occurs

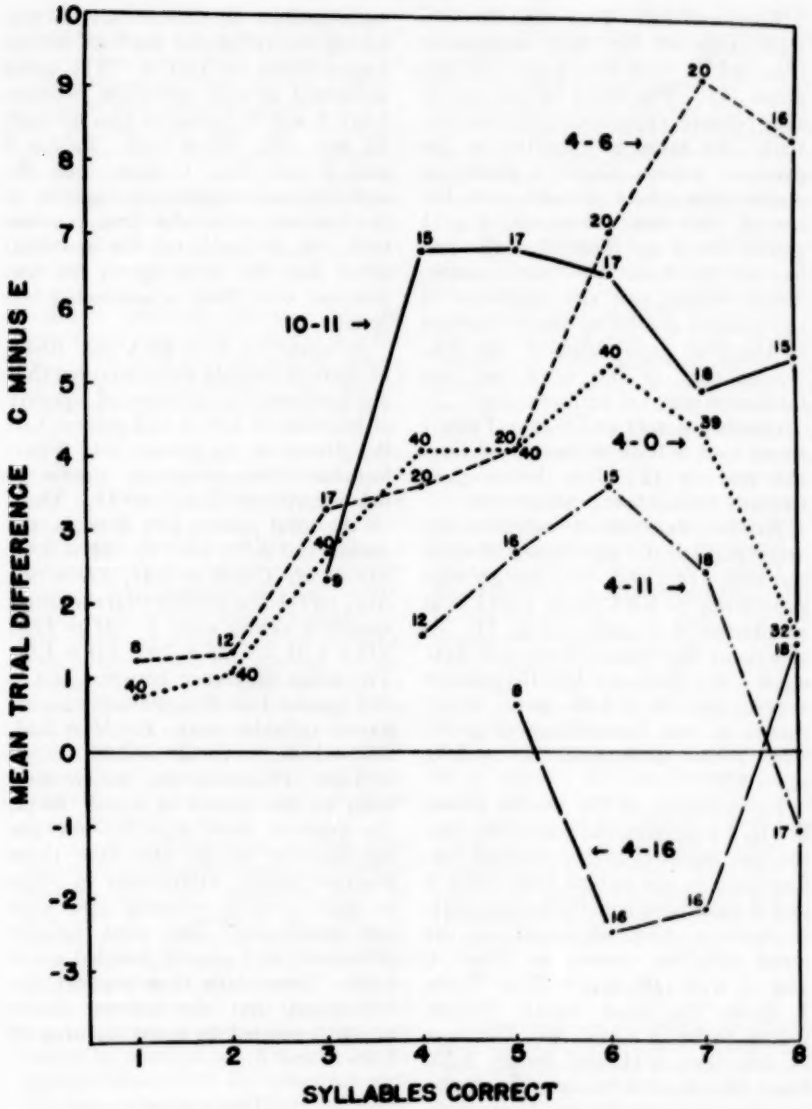

FIG. 1. Mean differences between control and experimental groups on postrest trials to successive syllable criteria.

between List A and List B. Due to the location of the rest, it is, of course, impossible for any effect of it to show on the first postrest trial.

Condition 4-6.—This condition is a replication of the first experiment reported in the first paper of this series (7). The effect of the rest in the present experiment is substantially the same as reported in the previous paper. Table 1 shows no appreciable effect immediately following the rest. However, Fig. 1 and Table 2 do show that the rest has an effect on later performance. These results and the similarity of the present results to those reported in the first paper support the conclusion that in this condition, rest facilitates postrest performance.

Conditions 4-11 and 4-16.—Table 1 shows that in both of these conditions the rest *Ss* (E) drop below their controls on the first postrest trial.

Neither experiment exhibits any other statistically significant effect of the rest. In Cond. 4-11 the greatest superiority of E4-11 over C4-11 is at a criterion of 6 correct (Fig. 1). At this point the mean difference is 3.60, $\sigma_m = 3.90$. In Cond. 4-16 the greatest superiority of C4-16 over E4-16 occurs at the same point (Fig. 1). The mean difference is -2.50, $\sigma_m = 2.99$.

Examination of the results shown in Fig. 1 suggested the possibility that the rest might facilitate postrest performance to the extent that Lists A and B have been equally learned prior to the rest. Over all conditions, the mean syllables correct on Trial 4, List A was 1.09 ($\sigma_m = .25$). Table 1 shows the most similar prerest values to be in Cond. 4-6, the next in 4-0, then 4-11 and finally 4-16. Since the general superiority of the experimental conditions over their respective controls (Fig. 1) shows the same ordering, Cond. 10-11 was run as a check on this possibility.

Condition 10-11.—In this condition only *Ss* were used whose total correct anticipations on prerest List B was within one syllable of the total correct anticipations on List A. The mean difference in each condition between Lists A and B (ignoring sign for each *S*) was .70, $SD = 1.22$. Tables 1 and 2 and Fig. 1 show that the experimental condition is superior to the control after the first postrest trial. As in Cond. 4-6, the beneficial effect does not show up on the first postrest trial, but is somewhat delayed.

In collecting data for Cond. 10-11, 18 pairs of records were obtained that did not meet the criterion of equality of learning of List A and prerest List B. However, on prerest List B performance they were very similar to the records of Cond. 10-11. Thus, on the total prerest List B trials, the means and *SD*'s were E—$M = 1.47$, $SD = .88$; C—$M = 1.42$, $SD = .95$. Also, on the last prerest trial the corresponding values were: E—$M = 2.61$, $SD = 1.61$; C—$M = 2.88$, $SD = 1.61$. The mean difference between List A and prerest List B in correctly anticipated syllables was, E—$M = 2.61$, $SD = 1.06$; C—$M = 2.33$, $SD = 1.11$. Following the rest, in contrast to the results of Cond. 10-11, the controls were superior but not significantly so on the first three postrest trials. Differences in trials to each syllable criterion also were not significant. The most reliable difference, at 5 correct, yielded a t of 1.10. These data thus support the hypothesis that the postrest facilitation is related to equal learning of Lists A and B.

Discussion

The important finding of the present experiments is that under the conditions employed, the rest aided further performance when the degree of learning of the two lists prior to the rest was approxi-

mately equal and hindered further performance when the prerest learning of the two lists was quite unequal. The other findings are that the rest may be beneficial with different degrees of prerest learning and that the beneficial effect apparently occurs shortly following the rest as well as later in the learning.

Of the many hypotheses that have been advanced to account for the beneficial effects of a rest, two alternatives appear most relevant to the present data. These are (*a*) that some inhibitory product generated by the memorizing process is reduced by the rest, or (*b*) that the rest acts to reduce interference between the two sets of responses so that either set may be more rapidly learned.

The first of these may be restated more specifically as the mental work decrement hypothesis discussed by Hovland and Kurtz (**4**), which assumes that engaging in mental work (thinking) on a problem generates some sort of inhibitory product that increases at a rate determined by the mental effort expended and that dissipates when *S* ceases working. This interpretation can account for the present data if it is assumed that by increasing the equality of learning of the two lists and thus increasing the difficulty, the *S*'s effort also is increased. However, a proper test of this hypothesis would seem to require that effort rather than difficulty be varied.

The second hypothesis assumes that the rest interval reduces the interference between the two competing sets of responses so that either set may be learned more efficiently after the rest. Since confusion between the two lists should be at a maximum when the two are of equal strength (**6**), the beneficial effect following a rest at this point also should be maximized. The data of the present experiments are consistent with this interpretation, but there are some serious problems. The implication of this idea is that during the rest, or as a result of the rest, generalized responses (i.e., the responses from the other list) tend to become weaker. Underwood and his associates (**9, 10, 11**) have recently developed an interpretation of distribution of practice effects that postulates just the opposite. They have assumed that when there is high interserial interference between lists, distribution of practice is effective because of the repeated extinction of erroneous responses that spontaneously recover during the rest intervals. By this argument a single rest should depress rather than facilitate performance. At the present time, the principle evidence supporting Underwood's hypothesis is that in learning lists of high interlist interference, the beneficial effects of distribution of practice come relatively late in learning, and with this sort of task, retention following distributed practice is superior to retention following massed practice (**11**). If this hypothesis is correct, the present data are most reasonably interpreted by the work decrement hypothesis. However, it should be pointed out that Hovland (**3**) found both reminiscence and facilitation of further performance under the conditions that Underwood and Richardson (**11**) suggest should minimize work inhibition. He obtained these results with *S*s who had previously practiced many similar lists.

Even though reminiscence has proved to be elusive, the beneficial effect of a single rest on further performance does not seem to be so difficult to produce. Either effect raises the same problem for Underwood's hypothesis. Therefore the possibility that a short rest may improve discrimination should not be rejected without more direct evidence.

Summary

An experiment was conducted that investigated the effect of a rest interval on performance when the rest followed varying amounts of original learning. A first list of paired associate nonsense syllables was rehearsed for four trials, and was followed without a break by a second list made up of the same stimuli but different responses. There were four experimental conditions that differed in the number of trials preceding the rest interval. The rest interval in the different conditions followed 0, 6, 11, or 16

trials of the second list. In each condition the 20 experimental *Ss* were matched against 20 control *Ss* that did not receive a rest. As measured by postrest learning, there was a slight beneficial effect of the rest in the 0-trial condition, a larger effect in the 6-trial condition and no positive effect when the rest was after 11 or 16 trials. There was a significant decrease in correct responses on the first postrest trial in these last two conditions.

As a result of these findings an additional condition was run. The *Ss* received 10 trials on the first list and 11 trials on the second list before a 2-min. rest was introduced in the experimental condition. For 20 pairs of *Ss* that learned the two lists to an approximately equal degree prior to the rest, the postrest learning in the experimental condition was superior to that in the control. For 18 pairs of *Ss* that did not meet this requirement, the rest did not facilitate performance. The conclusion is reached that the data are consistent with the hypothesis that the beneficial effect depends on the degree of competition between responses at the time of the rest. Two possible interpretations, a work inhibition hypothesis and a discrimination improvement hypothesis, are discussed.

References

1. Easley, H. The curve of forgetting and the distribution of practice. *J. educ. Psychol.*, 1937, **28**, 474–478.
2. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
3. Hovland, C. I. Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.
4. Hovland, C. I., & Kurtz, K. H. Experimental studies in rote-learning theory. IX. Influence of work-decrement factors on verbal learning. *J. exp. Psychol.*, 1951, **42**, 265–272.
5. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* (2nd ed.) New York: Longmans, Green, 1952.
6. Melton, A. W., & Irwin, J. McQ. The influence of the degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.
7. Riley, D. A. Reminiscence effects in paired-associate learning. *J. exp. Psychol.*, 1953, **45**, 232–238.
8. Riley, D. A. Further studies of reminiscence effects with variations in stimulus-response relationships. *J. exp. Psychol.*, 1954, **48**, 101–105.
9. Underwood, B. J. Studies of distributed practice: VI. The influence of rest interval activity in serial learning. *J. exp. Psychol.*, 1952, **43**, 329–340.
10. Underwood, B. J., & Goad, D. Studies of distributed practice: I. The influence of intra-list similarity in serial learning. *J. exp. Psychol.*, 1951, **42**, 125–134.
11. Underwood, B. J., & Richardson, J. Studies of distributed practice: XIII. Interlist interference and the retention of serial nonsense lists. *J. exp. Psychol.*, 1955, **50**, 39–46.

(Received March 19, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

EFFECT OF INTERTRIAL INTERVAL ON ACQUISITION AND EXTINCTION OF A RUNNING RESPONSE [1]

JOHN W. COTTON AND DONALD J. LEWIS

Northwestern University

Existing data on the effects of distribution of practice in the straight runway permit few general conclusions as to its effects. Increased intertrial intervals during acquisition have been variously reported as producing a concave downward function for acquisition running times (**2**), no significant changes (**3, 6, 7, 9**) and an almost significant increase in running time (**9**). Three studies (**6, 8, 9**) show no main effect of acquisition interval upon extinction, but one experiment (**3**) reports that massed acquisition yielded shorter extinction running times than did spaced acquisition. Massed extinction trials have led in some instances (**8, 9**) to shorter running times during extinction, in two instances (**4, 6**) to no effect, and in another case (**3**) to longer running times.

While some of the dissimilarities in the findings described above may be explained as resulting from differences in procedure in the different studies, enough disagreement remains to suggest the advisability of a single study in which a considerable range of intertrial intervals is studied during acquisition and extinction. The experiment reported below was designed to fill this need.

The present study employs a 4×4 factorial design in which each of four intervals (15 sec., 2 min., 8 min., and 16 min.) during acquisition is paired with each of the same four intervals during extinction in order to permit analysis of the effects of the acquisition variable, of the extinction variable, and of any interaction between these variables.

[1] This experiment was performed under a National Science Foundation grant (NSF G-1050) to the authors. We acknowledge the assistance of Rolland Metzger who participated in all phases of this study.

METHOD

Subjects.—The *S*s were 80 naive male albino rats purchased from the Sprague-Dawley Co. of Madison, Wisconsin. These 80 included replacements for 20 animals which were discarded because of apparatus failure or procedural mistakes, and 3 animals which were discarded because they failed to learn the running response. The *S*s were 55–67 days old at the beginning of experimentation, with a mean of approximately 60 days, and their mean weight was approximately 225 gm. Each *S* was assigned at random to one of 16 groups, subject to the restriction that the first 16 *S*s contained one member for each group, the next 16 one for each group, and so on, yielding a sample size of five for each group and balancing out any differences in ability of the early *S*s compared with *S*s run later in the experiment. Animals lost in a set of 16 were replaced at the end of the experiment.

Apparatus.—The straight runway employed by Lewis (**3**) and patterned after that of Sheffield (**7**) was used. The only lighting in the experimental room was provided by an overhead 100-w. bulb enclosed by a commerical shield. An automatic timer was activated when the starting-box door was raised and was stopped when *S* stepped on the goal-box door.

Procedure.—The procedure was divided into eight stages. The period of *deprivation and taming* extended for seven days in which each *S* was under 23-hr. food deprivation with 1 hr. of ad lib. feeding and a 5-min. period of handling each day. On Day 7 each *S* was placed in the closed goal box of the runway for 5 min. and required to eat three 45 mg. lab chow pellets from the food dish.

A *pretraining* session on Day 8 consisted of giving each *S* 10 rewarded trials through the complete runway, with a 15-sec. intertrial

interval.[3] One 45 mg. food pellet was given per trial. All *S*s were kept in individual living cages between trials and were allowed free access to water in those cages. On three occasions an *S* failed to complete a run within 120 sec. and was carried to the goal box and allowed to eat. Each was allowed 1 hr. of ad lib. feeding after the pretraining session was completed.

A *rescheduling of deprivation* period on Day 9 was introduced in order to make the average deprivation period equal for all *S*s despite the fact that the length of the experimental period ranged from 30 min. with the 15-sec. intertrial interval group to almost 8 hr. for the 16-min. group. On Day 9 each *S* was given 1 hr. ad lib. feeding beginning 23 hr. before the middle of the experimental session on Day 10.

A *training* session on Day 10 consisted of 30 rewarded trials for each *S* with one fourth of the 80 *S*s being assigned to each of the four acquisition intervals: 15 sec., 2 min., 8 min., and 16 min. Whenever *S* failed to reach the goal box within 50 sec., he was carried to the goal box and allowed to eat the pellet in order that the intertrial interval not be increased on slow runs. In these cases, of which there were only 11, 50 sec. was recorded as the running time for the trial. Each *S* was allowed 1 hr. of ad lib. feeding after the training session was completed.

A *rescheduling of deprivation* session on Day 11 involved providing 1-hr. ad lib. feeding 23 hr. before the middle of the ensuing day's experimental period in order to equate the average deprivation period for each group on Day 12.

An *extinction* session on Day 12 consisted of 30 nonrewarded trials, with 5 *S*s of each acquisition group being assigned a 15-sec. extinction intertrial interval, 5 to a 2-min. interval, 5 to an 8-min. interval, and 5 to a 16-min. interval. On any trial in which *S* failed to reach the goal box in 50 sec., he was removed from the apparatus and returned directly to his living cage for the remainder of the intertrial interval. A running time of 50 sec. was recorded for such trials.

A spontaneous recovery interval took place on Days 13 and 14, ending on Day 15 with a spontaneous recovery test with one nonrewarded trial under 22-hr. food deprivation. All *S*s were tested 72 hr. after the beginning of their extinction session. This resulted in some inequality from group to group in the interval from the end of extinction to the recovery test. Such inequality is not believed to have produced differential recovery since Ellson (1) has shown that spontaneous recovery, in a Skinner box at least, is substantially complete after 185 min.

Treatment of data.—Following Lewis (3), the median of each successive three running times during acquisition and extinction was determined for each *S*. The 10 acquisition medians for each *S* were averaged and converted to a common logarithm for analysis of acquisition effects. Similar treatment was given to extinction data. When trends were graphed, the mean of each two successive medians was determined and converted to a log value, these values being averaged over all members of a group.

Results

Acquisition.—Figure 1 summarizes the trend of running times with practice as a function of the acquisition interval. There is no indication that running time during acquisition depends upon the intertrial interval. An analysis of variance of the log mean median values for *S*s under the four acquisition conditions yielded an *F* of only .39.

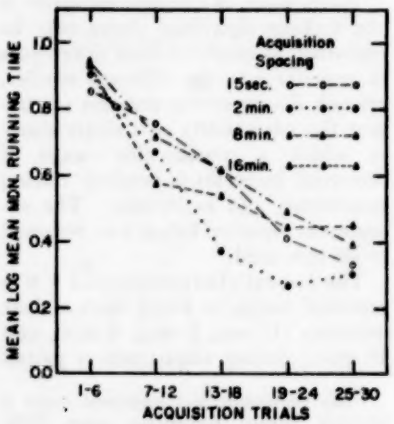

Fig. 1. The relationship between intertrial interval during acquisition and running time as acquisition training progresses.

[3] The 15-sec. group's intertrial interval is defined somewhat differently than for other groups. Each *S* spent 15 sec. in its living cage between trials, averaging about 1 min. from the start of one trial to the start of the next. In all other groups the intertrial interval is defined from the start of one trial to the start of the next. This procedural discrepancy is necessary because the rat's own running time is too variable to permit a fixed intertrial interval of less than 2 min.

Extinction.—Figures 2 and 3 present the trend of mean log mean median running times during extinction as a function of the acquisition and extinction intervals, respectively. Figure 2 shows no consistency of acquisition interval effects as extinction proceeds. Figure 3 shows that the 15-sec., 2-min., and 8-min. groups consistently arrange themselves from low running times to high in that order, whereas the 16-min. group is highly inconsistent in its position.

The over-all means for each of the 16 groups are presented in Table 1. This table indicates that running time increases markedly as the extinction interval increases from 15 sec. to 8 min. and decreases somewhat from 8 min. to 16 min.

Table 2, which summarizes analyses of variance and covariance upon the extinction data, shows that acquisition effects upon extinction are not significant whereas extinction effects are highly significant. The interaction between acquisition and extinction variables is also significant, apparently reflecting small discrepancies between the trend due to extinction interval following short acquisition intervals (15 sec. and 2 min.) and the corresponding trend following long acquisition intervals (8 min. and 16 min.) as shown in Table 1.

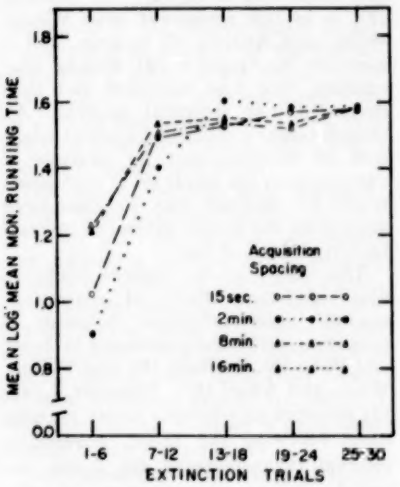

FIG. 2. The relationship between intertrial interval during acquisition and running time as extinction progresses.

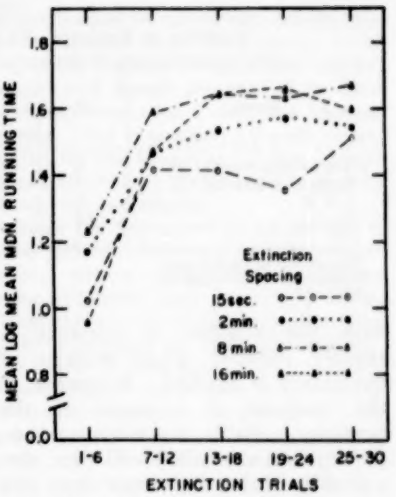

FIG. 3. The relationship between intertrial interval during extinction and running time as extinction progresses.

It will be noted in Table 2 that the significance of extinction effects shown

TABLE 1

EXTINCTION RUNNING TIMES AS A FUNCTION OF INTERTRIAL INTERVAL DURING ACQUISITION AND EXTINCTION
(Mean Log Mean Medians)

Extinction Interval	Acquisition Interval				
	15 sec.	2 min.	8 min.	16 min.	Mean
15 sec.	1.44	1.48	1.32	1.49	1.43
2 min.	1.43	1.48	1.60	1.53	1.51
8 min.	1.61	1.57	1.62	1.60	1.60
16 min.	1.55	1.53	1.59	1.54	1.55
Mean	1.51	1.52	1.53	1.54	1.52

TABLE 2

ANALYSES OF VARIANCE AND COVARIANCE OF EXTINCTION DATA

Source	*df*	Variance Analysis		Covariance Analysis	
		MS	*F*	*MS*	*F*
Acquisition interval (A)	3	.00476	.54	.00331	.37
Extinction interval (E)	3	.09888	11.17**	.08850	9.89**
A × E	9	.01842	2.08*	.01826	2.04*
Error	64	.00885		.00895†	

* Significant at the 5% level.
** Significant at the 1% level.
† *df* = 63.

with the analysis of variance is slightly reduced when analysis of covariance is applied. Reference to the analysis of variance of the predictor variable, log mean medians on acquisition trials 25–30 (not presented in Table 2), shows that this reduction resulted from the appearance of significant differences among extinction groups during acquisition ($F = 5.027$). The magnitude of reduction in significance by the covariance is small because the r within groups from acquisition to extinction is only .071, and the group differences during acquisition are not of exactly the same direction as in extinction.

Spontaneous recovery.—The log running times for the spontaneous recovery trial exhibit no main effects and no interaction effects of acquisition and extinction intertrial intervals. Since neither graphic nor statistical presentation of the spontaneous recovery data would modify the above statement, they are omitted from this report.

DISCUSSION

Two questions must now be considered: First, to what extent may these findings be integrated with results from previous studies of distribution of practice effects in a runway? Second, what theoretical conclusions are suggested by these results?

Four of the present results should be compared with previous data: (*a*) the absence of any main effect of acquisition intertrial interval upon behavior at any stage of experimentation; (*b*) the presence of main effects of the extinction interval upon extinction; (*c*) the presence during extinction of interaction effects between acquisition and extinction intervals; and (*d*) the absence of significant effects of either variable upon the spontaneous recovery trial.

The absence of significant group differences in acquisition data in this experiment supports the findings of Lewis (3), Scott and Wike (6), and Sheffield (7), is in fair agreement with Wilson, Weiss, and Amsel's (9) findings, and is contrary to Gagné's (2) finding that running time first increased and then decreased with intertrial interval. Although Gagné's statistical analysis might well be supplemented by analysis of variance over the whole set of acquisition trials, his findings may be considered anomalous but hardly subject to reversal by a reanalysis of data.

The failure of the present study to find significant effects of acquisition interval upon extinction behavior is consistent with results obtained by Scott and Wike (6), Sheffield (8), and Wilson, Weiss, and Amsel (9). However, Lewis (3) reported significantly faster running times following the 15-sec. acquisition interval than following the 2-min. interval. In an investigation of this apparent inconsistency between Lewis' and the present findings, Lewis' data and

the present data were reanalyzed after exclusion of all his groups except the four which employed 15-sec. and 2-min. intervals and 100% reinforcement and exclusion of all other groups except the corresponding four in this study. Analyses of variance and covariance then showed that in neither experiment are there significant effects of acquisition intertrial interval upon extinction behavior. Each study, however, shows some indication of a nonsignificant trend toward faster extinction response times following 15-sec. acquisition than following 2-min. acquisition intervals.

The appearance of significantly greater running times with the larger intertrial intervals during extinction supports the findings of Sheffield (**8**) and Wilson, Weiss, and Amsel (**9**) but contradicts the conclusions of Porter (**4**) and Scott and Wike (**6**), who found no evidence for any such differences. It may well be noted, however, that Porter's failure to find effects of intertrial interval may proceed from his exclusion from analysis of all extinction running times in which the animal failed to complete its run within 2 min. Porter notes in a footnote that such failures occurred 11 times out of a possible 60 in his 10-min. delay group, as compared to 6 times in his 0-delay group and 7 times in his 5-min. delay group. This trend is in harmony with the Cotton and Lewis finding that running time is slower with spaced extinction than with massed extinction.

The present findings also appear to diverge from previous results obtained by Lewis (**3**). He obtained no significant differences between groups over all 30 trials of extinction but found significantly smaller running times on the last 15 trials for the 2-min. group than for the 15-sec. group. The latter difference is not significant when data from Lewis' partially reinforced groups are excluded from the analysis to make it more comparable to the present study.

The interaction of the effects of acquisition and extinction intervals observed in the present study exhibit, for the 15-sec. and 16-min. groups only, the same magnification of extinction effects following massed rather than spaced training which Sheffield (**8**) found and Wilson, Weiss, and Amsel (**9**) verified in their first experiment for 15-sec. and 15-min. groups. No suggestion of such interaction for the 15-sec. and 2-min. groups appeared either in the present study or in Lewis' (**3**) experiment.

Since the findings of Lewis are somewhat different from those of the present paper, despite being obtained in the same apparatus, two procedural differences between the studies may have been possible sources of these divergencies. (*a*) Lewis' *S*s were kept on a deprivation schedule by being given a fixed amount of food daily whereas the present *S*s were given a fixed amount of time for eating daily. (*b*) Lewis' *S*s were removed after 2 min. without response during extinction trials whereas the present *S*s were removed after 30 sec. without response. The former difference could have produced greater drive in Lewis' *S*s as indicated by Ramond, Carlton, and McAllister's study (**5**). If drive and acquisition intertrial interval interact during extinction, Lewis' finding of significant acquisition interval effects during extinction and the findings of no such differences reported in this paper could conceivably be explained as a result of such interaction.

One final finding, that neither acquisition nor extinction spacing influences spontaneous recovery, supports Lewis' (**3**) findings of no main effects of these variables.

In the light of the foregoing comparisons of this experiment with earlier work, the best established fact about distribution of practice in a straight runway seems to be that massed extinction (15-sec. intertrial intervals) produces faster extinction trials than spaced (15- or 16-min. intertrial intervals) extinction. The present data make it clear that the relationship between degree of massing and running time is not linear or even monotonic. For this reason speculation as to the mechanism controlling this relationship

must be restrained. The Sheffield hypothesis that nonreinforcement produces a "frustration drive" which decays with increasing intertrial interval, thus producing faster runs with shorter intervals, remains basically acceptable but may require modification to account for the occurrence of peak running times with intertrial intervals of about 8 min.

Summary

Sixteen groups of five male albino rats were trained, extinguished, and given a spontaneous recovery trial in a 4-ft.-long straight runway, with intertrial interval as the only variable differentiating the groups. The 16 groups were arranged in a 4 × 4 design with 15-sec., 2-min., 8-min., and 16-min. intertrial intervals during acquisition and 15-sec., 2-min., 8-min., and 16-min. intertrial intervals during extinction.

The positive findings of this experiment are the existence of significant differences among extinction running times for the four extinction groups and a significant interaction of the acquisition and extinction variables upon extinction behavior. Fastest running times are shown under the 15-sec. extinction condition and slowest running times under the 8-min. condition. These findings are more or less consistent with Sheffield's frustration-drive hypothesis but raise the question as to why frustration drive should produce slower running with an 8-min. intertrial interval than with a 16-min. interval.

References

1. Ellson, D. G. Quantitative studies of the interaction of simple habits. I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23**, 339–358.
2. Gagné, R. M. The effect of spacing of trials on the acquisition and extinction of a conditioned operant response. *J. exp. Psychol.*, 1941, **29**, 201–216.
3. Lewis, D. J. Acquisition, extinction, and spontaneous recovery as a function of percentage of reinforcement and intertrial intervals. *J. exp. Psychol.*, 1956, **51**, 45–53.
4. Porter, J. M., Jr. Extinction of an acquired response as a function of the interval between successive non-rewarded trials. *J. comp. Psychol.*, 1938, **26**, 261–270.
5. Ramond, C. K., Carlton, P. I., & McAllister, W. R. Feeding method, body weight and performance in instrumental learning. *J. comp. physiol. Psychol.*, 1955, **48**, 294–298.
6. Scott, E. D., & Wike, E. L. The effect of partially delayed reinforcement and trial-distribution on the extinction of an instrumental response. *Amer. J. Psychol.*, 1956, **69**, 264–268.
7. Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.
8. Sheffield, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.
9. Wilson, W., Weiss, E. J., & Amsel, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, **50**, 51–65.

(Received June 18, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

THE ELECTRICAL PHOSPHENE THRESHOLD AS A MEASURE OF RETINAL INDUCTION AND VISUAL ORGANIZATION [1]

RICHARD M. MICHAELS

Naval Research Laboratory

In the past few years interest has been aroused in the use of the electrical phosphene, especially as a means of measuring some of the effects of visual stimulation on the retina. Much of the impetus for this interest comes from the work of Motokawa, who has published voluminously over the past decade (**1**). Very little research has been published either corroborating or contradicting Motokawa's findings. Gebhard (**2**) found none of the discontinuities in flicker phosphene thresholds that Motokawa reported. Michaels and Fox (**3**) were able to corroborate the threshold-time curves for yellow, green, and blue light reported by Motokawa (**4**). In view of the importance of the findings, it was felt that a more intensive attempt be made to test some of Motokawa's conclusions. From a perceptual standpoint one of the more promising experiments was that on spatial induction, first described in 1949 (**5**). This effect is essentially a spreading of excitation in the retina about the borders of a projected image. The excitation decreases monotonically from the borders of the figure, the rate of decrease depending upon the brightness of the figure (**5**), its spatial characteristics (**6**), and distortion due to the retina itself (**6**).

Testing the induction effect was undertaken for two reasons. First, Motokawa suggested that induction was a more clear-cut effect than the zeta-time curves. Thus testing spatial induction seemed to offer a greater opportunity for corroboration of the electrical stimulation technique. Second, spatial induction offers an explanation for certain visual perceptual phenomena. If interaction occurs among the fields induced by a number of discrete figures, then induction might represent a physiological correlate of certain Gestalt properties such as grouping or continuity. Thus the present study was an attempt to confirm the spatial induction effect and simultaneously to test the hypothesis that the induced fields about two dots interact and combine thus giving a sensation of continuity.

METHOD

Apparatus.—The apparatus may be described in three parts. The first is the visual stimulator which is shown schematically in Fig. 1. The inducing stimuli were two yellow dots projected onto the viewing screen by reflection off the reflecting plane of the half-silvered mirror. The

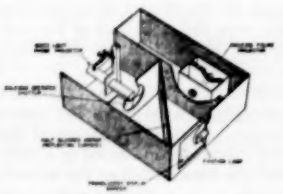

FIG. 1. Schematic view of the apparatus for presenting the visual stimuli.

[1] The substance of this report was presented at the American Psychological Association meeting, September 1955. Acknowledgment is gratefully extended to Mr. W. J. Finney of this Laboratory and Dr. B. H. Fox of the George Washington University for their support and assistance in matters electronic and psychologic.

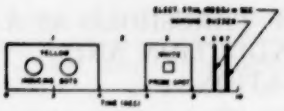

FIG. 2. The loci of measurement of spatial induction.

projected image gave two dots 7 mm. in diameter and 7 mm. apart. The brightness of these dots at a viewing distance of 50 in. was 3.0 ft.-Lamberts. The white probe light source projected an image through the half-silvered mirror giving a focused image that was 1.0 mm. wide and 1.2 mm. long and having a brightness of 1.3 ft.-Lamberts at 50 in. This probe projector was mounted on a microscope stage so that it could be moved in two dimensions about the stimulus dots and its position set to the nearest .1 mm. A small, red fixation light was placed 2.5 in. from the yellow inducing figures and aligned with them. Finally the projection screen was a 6 × 6 in. piece of translucent tracing paper.

Current was used as the independent variable since it had been shown by Michaels and Fox (3) that current was the significant electrical measure. A simple constant current source was constructed that minimized the effect of changing skin resistance on the amount of current passing through the eye. The source had an accuracy of 2%, and variability due to changes in skin resistance was approximately .2%.

The timing circuit was a set of specially modified pulse-operated relays (Burroughs Model 1801). These fired in sequence, and would cycle continuously until the operating pulse was externally interrupted. The duration of closure of each of the relays was adjusted using a chronoscope and it was found that the duration of closure could be set to a .1% accuracy.

Standard 10-mm. EEG discs were used and were chlorided just prior to testing. The electrodes were mounted in two spring-loaded lucite blocks on an adjustable headband. Thus pressure could be varied as well as positioning the block on the head. The positive electrode was placed on the frontal bone just above the brow and over the nasal half of the eye, while the negative was placed on the homolateral temple.

Procedure.—The use of the phosphene threshold to measure spatial induction depends on three factors. First, if the electrical threshold to a white stimulus is measured as a function of time after the presentation of the white light, a characteristic curve is obtained with a maximum at 2 sec. (4). Second, if a stimulus of some hue precedes this white light it induces a marked change in the characteristic white-light curve (5). Third, the greatest change in this white-light curve occurs at that time which is the maximum of the time curve of the hue used, provided the white stimulus is projected on an area outside the border of the projected colored stimulus (5). In this study yellow light was the inducing stimulus used to modify the white-light curve. Figure 2 shows the paradigm of the experiment with the attendant times. The warning buzzer occurred .5 sec. before the electrical stimulus.

With the electrodes affixed, *S* was dark adapted for a period of at least 20 min. At the end of that time the phosphene threshold in the dark was obtained since all changes of threshold in subsequent conditions were relative to this one. Then the threshold to the white probe light was obtained, the electrical stimulation occurring 1500 msec. after the white. This time was the maximum of the threshold-vs.-time curve for yellow light and was intended to maximize the electrical threshold change to the white light due to the inducing effect of the yellow. After this threshold was obtained, the yellow inducing dots were presented prior to the white light. The sequence of events was then as shown in Fig. 2. The phosphene threshold was obtained to the white probe when it was located at various points about the border of the inducing figure. The difference in threshold of white relative to the dark threshold before and after the yellow stimulus was the measure of the induction effect. That is, induction $= (I_0/I_{w'} - 1)\ 100 - (I_0/I_w - 1)\ 100$, where I_0 is the dark threshold, and I_w and $I_{w'}$ are the white-light threshold before and after presentation of the inducing figures, respectively.

In order to construct the field of induction about the two dots, two conditions were imposed. First, to reduce the number of measurements needed it was desirable to test for symmetry about the pattern. This was done by obtaining thresholds at selected symmetrical points about the pattern. The second condition was that, with symmetry, a minimum of five induction-vs.-distance curves be obtained in one quadrant. This quadrant was defined by the vertical

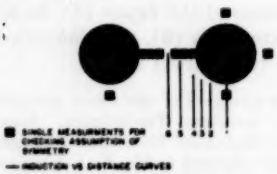

FIG. 3. The time sequence for presenting the stimuli.

diameter of one of the dots and the vertical line, median to the two dots. All measurements were made on lines perpendicular to the line of regard. Figure 3 shows the conditions tested.

Two female *Ss* were used, neither of whom had any visual abnormalities. Testing was done one day a week for each, usually lasting for 3 hr. In this time 5 to 12 points were obtained.

Four different procedures for obtaining thresholds were investigated: (*a*) method of limits; (*b*) method of constant stimulus; (*c*) a method of successive comparisons, as recommended by Motokawa (7), in which *S* judges which of two successive stimulations elicited a phosphene when one of the two was well below threshold; and (*d*) a method of successive approximations. The latter was ultimately followed and was found to be the most economical. A threshold was obtained by first presenting *S* with a pulse well above threshold (10–30 μamp.) and then reducing the succeeding pulses in 4 μamp. steps. When the range in which threshold lay was thus determined, the current was first raised above threshold, on the second succeeding one reduced below, the third then above but lower than the previous supra-threshold stimulus, the fourth below but higher than the previous sub-threshold stimulus. This process of successive approximations was continued until the threshold was bracketed to within 2–4 μamp. Presentation of these two stimuli either alternately or in random or counterbalanced order was continued until three successive or four out of five positive and negative responses were obtained respectively to the supra- and subthreshold stimuli. With this criterion, the threshold was taken as the median current between the two, and it was possible to estimate the thresholds to within ± 3%.

Results

The results which follow were obtained over a three-month period from January through March of 1955. Testing with these *Ss* began, however, in August of 1954. Over four months training was necessary for these *Ss* to establish a stable discriminatory frame of reference.

TABLE 1

Distribution of the Absolute Dark Threshold in μamperes

S	Range	Mean	*SD*	*N*	95% Conf. Interval
P. K.	50–55	52.63	1.65	8	± 1.38
D. S.	61–63	61.72	.70	7	± .65

TABLE 2

Effect on Relative Threshold of Changing Absolute Threshold

Thresholds (in μamp.)		Relative Threshold
Dark	White	
114	94	21
98	84	17
74	63	17
62	53	17

There are essentially four classes of results that are of interest. The first concerns the stability of the threshold in absolute terms. The measure chosen for this was the dark threshold since this was dependent only on variability within *S*. Table 1 contains the pertinent statistics.

The second class of results deals with the stability of the relative threshold. Here information was obtained from only one *S* because the other showed no stability until her thresholds reached their final levels. With the first-mentioned *S* the relative change was measured between the dark threshold and the white-light threshold when the yellow light was absent. Table 2 shows the relative change that occurred over a four-week period. The relative threshold change was sufficiently small with fluctuations in absolute threshold to make unnecessary repeated measurements of the white-light thresholds. This fact coupled with the ultimate stability of the absolute threshold made it unnecessary to measure either of these two thresholds, and so in later induction measurements they were not made, except as a check at irregular intervals.

The third class of results deals with

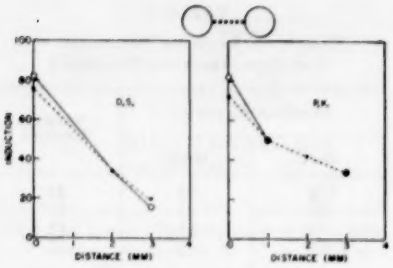

FIG. 4. Induction effect at symmetrical points between the two figures.

the hypothesis of symmetry about the yellow-inducing figures. Figure 4 is a plot of the induction effect for the points between the two figures as shown in Fig. 3. The curve for P. K. was measured along the horizontal line connecting the diameters of the two dots. However, for D. S. the data were inadvertently taken along a horizontal line displaced by 1 mm. from the line connecting the diameters. The points, of course, are still symmetrical. For the remainder of the points in Fig. 3 used to check symmetry, it was found that the induction effect was the same for all. Thus the assumption of symmetry appeared reasonable and all the subsequent curves were taken in the quadrant as shown in Fig. 3.

The fourth class of results concern the spatial induction phenomenon. Figure 5 is the induction-vs.-distance curves which were fitted by eye. These points represent the combined results of two or more measurements along the five lines with the exception of the 2 mm. case for D. S. In general, variability increased as measurements were made more intermediate between the inducing figures and also there was a decrease in the stability of individual thresholds in this region. Finally, from Fig. 5 it was possible to construct the field of iso-induction contours about the figures. Figure 6 shows these fields of induction for each of the *Ss*. The points noted on the figure mark the area where the data were taken. The remainder of the field was constructed on the assumption of symmetry. These fields were drawn by interpolation from Fig. 5 and the accuracy of the contours is naturally dependent upon the accuracy of the curve fitting which at best is only approximate.

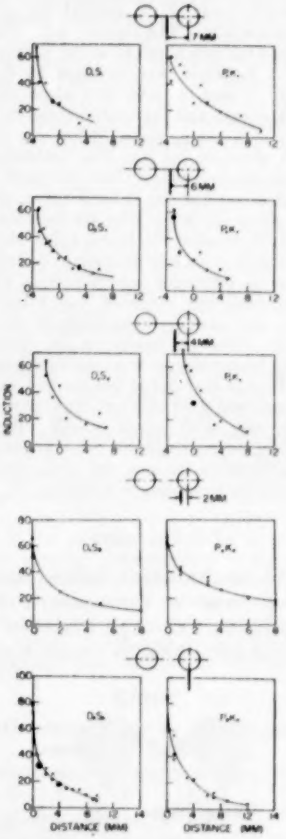

FIG. 5. Change of induction with distance. (Coordinates are with reference to the point tangent to the vertical diameter.)

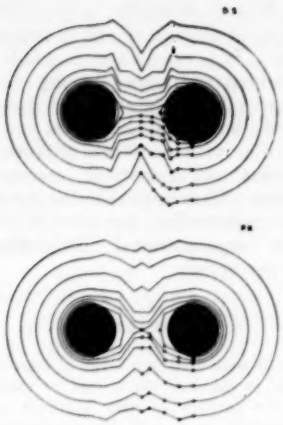

FIG. 6. Iso-induction contours about two yellow dots.

DISCUSSION

The electrical phosphene threshold presents a definite measurement problem. It is very sensitive to slight changes in the set of the *S* and methods of obtaining it. As was pointed out, over four months were required to obtain reliable thresholds. In early stages of training it was found that *S* tends to respond to the presence of a visual sensation. Both *S*s tended to discriminate a diffuse brightening as a phosphene. Using this as a criterion, however, led to dark thresholds for one *S* in the region of 100–150 μamp. and trial-to-trial variability of the order of 10–20%. In the other *S* (P. K.) there was a continuing adaptation, an intra- and intersession increase in threshold. This *S* in the first two sessions gave a stable threshold using the successive approximation technique of 57 μamp. At the end of four months it was necessary to apply 300 μamp. to elicit a consistent positive response, and no threshold to the established criterion could be obtained. It seems reasonable to conclude that these shifts and inconsistencies were due to an unstable frame of reference. On this assumption the *S*s were forced to scale the brightness of the phosphene; were given catch stimuli approximately one-third of the time; and were forced to make choices between successive stimuli; and it was at this time that the warning buzzer was added. Whatever other effects these techniques may have had they did motivate the *S*s to search for different cues and criteria for judgment. Within a month thresholds were at or close to their ultimate levels.

The method of successive approximations is not a standard psychophysical method and suffers from a lack of validation common to the more classical techniques of measurement. It was developed out of the need for an economical measure for which a reasonable criterion could be established. As was noted, three other methods were tried. The method of limits suffers from three limitations: (*a*) there is no assurance that equal current steps are equal psychophysical steps; (*b*) a large number of trials are required for reasonable precision, therefore taking a great deal of time; (*c*) there are adaptation effects to stimuli close to threshold. This last was the most consistent feature of these thresholds. In a run of three or four stimuli which are the same or decrease in small steps, the third or fourth tend to elicit a negative response if they are slightly above threshold. Under such conditions it is necessary to raise the stimulus considerably to obtain a positive response. The same adaptation effect also occurs for stimuli slightly below threshold so that *S* will respond positively. That this is an adaptation effect is evident from the fact that a catch stimulus will usually elicit a positive response in the first case and in the second case a suprathreshold stimulus will usually elicit a negative response. Actually this reversal tendency was occasionally stable enough to meet the threshold criterion, but when measurement was repeated the same thresholds were obtained without reversal. It seems reasonable that these reversals represent fluctuations or adaptation in *S*s' frames of reference rather than some physiological property of the eye. The *S*s used as a criterion more of a recognition of a change in state rather than an actual perception of light. Thus their reference was related to the stimulus series rather than some invariant or absolute external criterion. Near threshold, then, the discrimination reduces to a perception of difference between contiguous stimuli. Runs of the same stimulus which offer little change would tend to minimize the probability of a perception of difference and thus make the judgment unstable. Such a situation seems consistent with the results reported by Verplanck, Collier, and

Cotton (8) on the nonindependence of responses of visual threshold measurements.

The constant methods suffer from the same disadvantages as the method of limits. Motokawa's method of successive comparison has the first and third limitation mentioned for the method of limits. Furthermore, *S* is allowed to respond with "I don't know." With well-trained *S*s this may be adequate but there are cogent arguments against its use.

The method of successive approximation seems to avoid the limitations found with the other methods. If equal current steps are not equal psychophysical steps the procedure tends to bias the accepted threshold away from the current median to the two terminal values. Since the terminal values must elicit a recognition probability of between 0–25% or 75–100% the bias cannot place the threshold beyond these terminal values. The greatest limitation of the method would appear to be its restriction in range of recognition since thresholds are obtained by interpolation from only two points of some transition curve. There is, therefore, no practical measure of variability for any one threshold determination.

In general the results offer verification of the spatial induction effect. A major difference is that the data have considerably more variability than is indicated in Motokawa's results. The reasons for the difference are unknown; however, the complexity of the discrimination as found in this study would lead to an expectation of high variability. Also, of course, testing was done only once a week and three months were required to obtain the data presented here.

Finally from Fig. 6 it may be seen that there is a spread of excitation on the retina, developed as a direct consequence of the presence of a projected image. Where there are two figures, the fields derived for each interact to bring about unification of the discrete elements. This type of organization has implications in visual perception, for it allows an understanding of some properties emphasized by Gestalt experiments such as grouping and continuity. The results do in fact demonstrate a physiological correlate of continuity and a retinal one at that. It is possible to conclude that a perceived continuity among discrete elements is due to the spread of excitation about the borders of the figures and their interaction with each other, arousing a perception of continuity among the elements. Thus spatial induction offers partial explanation for certain of the simpler organizational properties of visual perception.

Summary

Two *S*s were used in a series of experiments designed to measure the spread of excitation about the borders of two identical circles. Measurements were made by determining the electrical phosphene threshold, as the means of measuring the change in the characteristic curve of white light caused by prior stimulation with two yellow dots. Thresholds to the white light were obtained as it was placed in various positions about the border of the inducing figure, thus permitting the construction of the iso-induction field about the yellow dots.

The results and conclusions were:

1. Reasonably stable thresholds were obtained after five months of training using a combination of the method of limits and a series of successive approximations to the limens from above and below threshold.

2. The field about the two inducing dots was found to be symmetrical, and to decrease monotonically with distance from the border of the figure, thus verifying Motokawa's findings.

3. The variability of the induction curves tended to increase for regions of greater interaction between the induced fields.

4. It was suggested that the Gestalt properties of perceptual grouping have their bases in the structure of the induced fields which organizes the discrete elements into a unified figure. Thus, spread of excitation at the retinal level may be a source of the organizational property in such Gestalt phenomena.

References

1. Gebhard, J. W. Motokawa's studies on the electric excitation of the human eye. *Psychol. Bull.*, 1953, **50**, 73–111.
2. Gebhard, J. W. Thresholds of the human eye for electric stimulation by different wave forms. *J. exp. Psychol.*, 1952, **44**, 132–140.

3. Michaels, R. M., & Fox, B. H. The threshold of electrical phosphenes as a function of the time interval after stimulation with various hues. Paper read at Eastern Psychol. Ass., Atlantic City, April, 1953.
4. Motokawa, K. Retinal processes and their role in color vision. *J. Neurophysiol.*, 1949, **12**, 289–303.
5. Motokawa, K. Physiological induction in human retina as basis of color and brightness contrast. *J. Neurophysiol.*, 1949, **12**, 475–488.
6. Motokawa, K. Field of retinal induction and optical illusion. *J. Neurophysiol.*, 1950, **13**, 414–426.
7. Motokawa, K. Some remarks on measurement of electrical excitability of the eye. *Tohoku J. exp. Med.*, 1951, **54**, 385–392.
8. Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, **44**, 273–282.

(Received June 11, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

LEARNING THE STATISTICAL PROPERTIES OF THE INPUT IN PURSUIT TRACKING[1]

E. C. POULTON

Applied Psychology Research Unit, Cambridge, England

In tracking sine waves without advance vision, *S* can use two sources of information in order to predict the input of the immediate future (**1**). First, it is possible to predict the change in the position and rate of movement of the input from its present rate and acceleration, for there are no sudden discontinuities. And second, certain statistical properties of the input can be learned, such as its mean and maximum amplitude. It is not normally possible to determine the relative importance of these two sources of information, or even to be sure that both are in fact used.

In contrast, in the present tracking task the courses consisted of constant slopes which were separated by sudden discontinuities in direction. It was the positions of these sudden discontinuities which had to be predicted if tracking was to be accurate. They could not be predicted from the input rate or acceleration, for between the discontinuities the input rate remained constant and the input acceleration was zero. Prediction therefore had to be based upon the statistical properties of the input. The use of this source of information could thus be studied in the absence of the use of the rate and acceleration of the input. In this respect the task corresponds to step-function tracking (**2**).

The present experiment also compares the effectiveness of visual information about a course obtained before tracking, with (*a*) advance visual information obtained while tracking, and (*b*) information about the future course deduced while tracking from previous systematic trends.

METHOD

Courses were drawn in India ink on white cardboard. Each course consisted of three W's or M's side by side with no gap between them. The lines were about .7 mm. wide, and were inclined at 13.5° to the vertical. The sum of the lengths of the 12 lines comprising a course was 25.0 in. The maximum amplitude or height was 4.0 in. and the course covered 6.0 in. of the cardboard in a horizontal direction.

Eight courses were patterned. The patterns were: six V's all of equal size; three large V's alternating with three small V's; a large, medium, and small V in that order, repeated once; the same pattern upside down, each half starting small and ending large; three M's with each of the 12 lines smaller than the previous line; the same pattern upside down, with successive lines larger than the previous line; three W's with the lines first getting larger and then smaller again; and three W's with the lines first getting smaller and then larger again.

The other eight courses were as far as possible random. For each course 12 numbers were selected at random from a table of random numbers, with the restriction that no number could be less than 2 or greater than 16, and that the total had to add up to 96. This series of numbers gave the relative sizes of the lines which comprised the course. Four of the courses looked like three irregular W's side by side. The other four courses ought to have looked like three irregular M's side by side, but by an oversight one course had only 11 sides, although it was the same length as the others.

Tracing paper was fixed over each course, and the course had to be traced with a pencil from left to right. A clear celluloid disc 25.0 cm. in diameter was fixed to the pencil 2.25 cm. above the paper. A circular paper mask of the

[1] This work was under the direction of Prof. Sir Frederic Bartlett, F.R.S., and Dr. N. H. Mackworth. The *S*s were supplied by the Royal Navy. Financial support from the British Medical Research Council is also gratefully acknowledged.

same diameter could be fitted over the celluloid. The mask was white underneath and black above. It had a circular hole in the center 2.0 cm. in diameter. The pencil itself blocked the middle .7 cm. of this circular window. One version of the mask, the half mask, had a 90° sector cut out to within 2.5 cm. of the central window. This open sector was placed on the left, so that it only revealed the part of the course which had already been traced, together with the corresponding edges of the cardboard. The course was illuminated from the right and left sides by two shaded 60-w. lamps on each side.

There were four conditions (see Table 1). In the condition of knowledge and no mask, the clear celluloid disc was attached to the pencil without a mask. The *E* placed the pencil at the start of the course, and *S* could start tracing whenever he was ready. In the conditions of knowledge and half or full mask, *S* was allowed to see the course for as long as he liked before the pencil was placed in position. The time taken was noted by *E* from a stop watch. The *E* then placed the pencil in position with the mask already attached to it. In the condition of no knowledge and full mask, *S* had to trace with the full mask without having seen the course previously. Each *S* traced two patterned and two random courses under each of the four conditions. The design used a 16×16 latin square in which both conditions and types of course were varied systematically.

The *S* was told that half the courses were patterned and half random, and that the order of the two types was random. He had to take about 10 sec. to trace each course. He was timed with the stop watch and told afterwards how long he had actually taken in relation to the target of 10 sec.

There was a blank space of 2.0 in. at the end of each course. When *S* reached the end he had to extrapolate the course in this space, with the mask if present still in position. There was no time limit for this extrapolation. He was then asked what this extrapolation was intended to represent. If a mask had been used for the tracing, he was also asked to report the extent to which he had been able to predict the nature of the course ahead of his pencil.

There was a practice immediately before the experiment. In this *S* first learned to hold the pencil vertical while tracing (or the celluloid disc hit the tracing paper). He was then introduced to the experimental conditions, and traced patterned and random courses. The practice and experiment together took about an hour. Each *S* performed the experiment once only.

Subjects.—The 16 *S*s were all enlisted men in the British Royal Navy, aged between 18 and 25 yr.

Scoring and calculations.—Two measures of error were taken: (*a*) the vertical distance between each of the 11 corners in a course and the corresponding position at which *S* changed direction, and (*b*) the horizontal distance between the course and the tracing, at the midpoint of each of the 12 lines in the course. The mean constant error (CE) at the corners was calculated by subtracting undershoots from overshoots. For the mean error the distances were summed regardless of sign.

Lines approaching corners were classified as "long" if they were greater than 2.1 in. in length, and "short" if they were 2.1 in. or less. A corner was classified as "central" if the line approaching it stopped short of the center of the range of possible positions for corners, or if it extended not more than .8 in. beyond the center of the range. Otherwise the corner was classified as "peripheral." For the patterned courses, the position of a corner was taken to be "predictable" if it was the third or beyond in a systematic trend. Otherwise it was considered to be "unpredictable."

The first corner of each pattern has been excluded from Tables 2 and 4, because there was no previous line, only an approach line. Two other corners, one from a patterned course and one from a random course, have been excluded from Table 2 only, because the lengths of the previous and approach lines did not fit the classification used in the table.

The significance of the difference of each pair of means was estimated both against the variability between *S*s, and against the variability between courses or parts of courses. The less significant of the two levels was the one taken. If anything, levels of significance have been underestimated as compared with more conventional procedures. Two-tailed tests have been used except where stated. For measures of time, medians have been taken because the distributions were skewed (3). The significance of differences in time were calculated from the medians for *S*s and courses.

Results

When *S*s were allowed to study the courses for as long as they liked before tracing them with a mask, the median times taken for study were 6.0 sec. for patterned courses and 9.0 sec. for random courses ($P = .01$).

The main results of the tracing are

TABLE 1

MAIN RESULTS

Condition	Median Total Time (Sec.)*		Mean CE, per Corner (Mm.)†		Mean Error per Corner (Mm.)‡		Mean Error per Midpoint (Mm.)¶	
	Pattern	Random	Pattern	Random	Pattern	Random	Pattern	Random
1. Knowledge, no mask	8.9	8.9	− .08	− .07	1.17	1.49	1.11	1.16
2. Knowledge, half mask	12.9	13.5	.71	.81	2.09	2.45	1.26	1.27
3. Knowledge, full mask	12.4	14.0	.68	1.13	2.31	2.21	1.26	1.25
4. No knowledge, full mask	14.0	15.2	.90	1.49	2.25	2.69	1.21	1.18

* Condition 3, $P = .05$ for Pattern vs. Random. For Pattern and Random combined, Cond. 1 and 4 different from all others at .02 level or better.
† Minus values indicate a tendency to undershoot. For Random alone, $P = .05$ (one-tailed test) for Cond. 2 vs. Cond. 4. For Pattern and Random combined, Cond. 1 different from all others at .01 level or better.
‡ For Pattern and Random combined, Cond. 1 different from all others at .01 level or better.
¶ None of these differences is significant.

given in Table 1. With the random courses the increase in CE between knowledge and half mask and knowledge and full mask was due to overshooting in the first halves of the random courses with knowledge and full mask. In this condition the CE for the first halves was 2.5 times larger than the CE for the second halves ($P = .01$).

In about 10% of the trials with a mask, the error was so large at some point that *S* could not see the line he was supposed to be tracing. The condition of no knowledge accounted for about half of these cases. In about one-third there was some delay before *S* was able to find his way back to the line. These delays slightly increased the median tracing times of the conditions with masks. Wandering too far from the line also sometimes affected the error as measured at the midpoints of the lines in the courses (see *Scoring and calculations*). But wandering too far did not affect the error as measured at the corners.

Tables 2 and 3 show the principal determinants of the size of the CE at the corners. The distribution of corners in Table 3 is independent of the distribution in Table 2. It was the later parts of the patterned courses which tended to be predictable. However, order within the patterned courses was not found to make any difference when a group of early corners was equated for predictability with a group of late corners.

When *S* was able to see a patterned course before or while tracing it, he was able to extrapolate it more or less correctly at the end of the tracing in about three-quarters of the trials. When he was not able to see the patterned course he was tracing, his extrapolation was only more or less correct in about one-third of the trials. In another third of the trials *S* claimed that he had correctly grasped the nature of the pattern. But he also claimed to have grasped the nature of the "pattern" under these conditions in a quarter of the trials with the random courses!

DISCUSSION

Learning what to expect when tracking with the full mask.—Table 2 shows the relationships between the CE at the corners and the nature of the part of the course being tracked. Each of the three variables—position of corner, length of approach line, and length of previous line—produced a significant difference in

TABLE 2

EFFECT OF RECENT PART OF COURSE

Condition	Mean Constant Error per Corner in Mm.											
	Central Corner					Peripheral Corner						
	Previous Line ≤ 2.1 In.		Previous Line > 2.1 In.			Previous Line ≤ 2.1 In.				Previous Line > 2.1 In.		
	Approach Line ≤ 2.1 In.		Approach Line ≤ 2.1 In.		Approach Line >2.1 In.	Approach Line ≤ 2.1 In.		Approach Line > 2.1 In.		Approach Line > 2.1 In.		
	Pattern	Random	Pattern	Random	Random	Pattern	Random	Pattern	Random	Pattern	Random	
1. Knowledge, no mask	.35	.03	.31	−.11	.06	.00	.37	−.38	.07	−.35	−.52	
2. Knowledge, half mask	.88	1.10	.63	1.74	1.88	.56	.25	1.19	.30	.28	.23	
3. Knowledge, full mask	1.75	.89	1.78	1.65	1.25	−.60*	.73	1.47†	1.17	−.04	.89	
4. No knowledge, full mask	1.55	1.28	2.69	4.48‡	1.75	.25*	.88	.41	.93	.44	−.09‡	
N for corners	17	16	8	14	4	13	15	8	15	33	14	

Note.—Minus values indicate a tendency to undershoot.
* Central Corner-Peripheral Corner $P = .02$ or better.
† Approach Line ≤ 2.1 In.-Approach Line > 2.1 In. $P = .05$.
‡ Previous Line ≤ 2.1 In.-Previous Line > 2.1 In. $P = .05$ (one-tailed test) or better.

CE under at least one condition with the full mask, when the other two variables were held constant. Table 2 also gives the frequency with which each particular combination of variables occurred in the experimental courses.

The following evidence in Table 2 suggests that, when tracking with the full mask, the CE tended to be smaller with probable combinations of course variables than with less probable combinations; or, to put it more simply, that *S* to some extent learned what to expect. (*a*) With both short approach and previous lines and with long, the CE tended to be smaller with peripheral corners than with central corners. Peripheral corners were 1.7 times more frequent than central corners in both the practice and experimental courses. (*b*) For a given position of corner, the two combinations short (previous line)-short (approach line) and long-long tended to give smaller CE's than the two combinations short-long and long-short. These results again correspond to the relative frequencies of the combinations. These frequencies are given in Table 4 for the practice and experimental courses combined.

Table 3 shows that when the *S* tracked a patterned course with the full mask, the CE at a corner was smaller when its position could be predicted from the previous part of the pattern than when its position could not be predicted in this way. This effect was independent of the effect of frequency discussed

TABLE 3

EFFECT OF COURSE PATTERN

Condition	Mean CE per Corner in Mm. When Position is	
	Predictable	Not Predictable
Knowledge, no mask	.21	.04
Knowledge, half mask	.49	.97
Knowledge, full mask	.23	1.10*†
No knowledge, full mask	.45	1.33*
N for corners	43	45

* Predictable − Not Predictable $P = .05$.
† No knowledge Predictable − Knowledge Not Predictable $P = .05$ with a one-tailed test.

TABLE 4

FREQUENCY OF COMBINATIONS OF LENGTH OF PREVIOUS LINE WITH LENGTH OF APPROACH LINE

Previous Line	Approach Line	
	⩽ 2.1 in.	> 2.1 in.
⩽ 2.1 in.	77	30
> 2.1 in.	31	61

above. It can therefore be concluded from *S*'s records that he learned to some extent what to expect in each particular patterned course and what to expect in any course.

Visual information before tracking was relatively ineffective.—As compared with the condition of no mask, Table 1 shows that the greatest increase in tracking time and in error at the corners occurred when a mask was introduced. There was only a relatively small further increase when *S* was not allowed to see the course before he started to track. Table 1 also shows that it made relatively little difference whether a half mask was used, which allowed *S* to see behind, or a full mask which did not. Thus the mask produced most of its detrimental effect by restricting advance vision, and visual information about the course obtained before tracking was unable to compensate for this.

Table 3 shows that, in tracking a patterned course with the full mask, the CE was larger at unpredictable corners when knowledge was acquired visually before tracking than at predictable corners when no knowledge was acquired before tracking. This again shows the relative ineffectiveness of visual information acquired before tracking.

These differences in effectiveness may be due, at least in part, to the nature of the coding. The information obtained before tracking with a mask was purely visual, whereas the information obtained while tracking without a mask and while tracking the predictable courses with a mask was necessarily mixed visual and kinaesthetic. Information in this latter form is probably easier to use in tracking.

SUMMARY

An experiment on a form of pursuit tracking was designed (*a*) to determine *S*'s ability to learn, and use his knowledge of, the statistical properties of the input, and (*b*) to compare the effectiveness of visual information about a course acquired before tracking, with visual-kinaesthetic information acquired while tracking.

The *S* had to trace with a pencil courses which consisted of constant slopes separated by sudden discontinuities in direction, to meet a time criterion. Half the courses were patterned, in that they contained systematic trends which could be used in prediction. The other courses were more or less random. In three conditions vision was restricted, by a mask attached to the pencil, to the part of the course near the tip of the pencil. In one of these conditions the part of the course which had already been traced could also be seen, and in this and one other condition the course could be studied visually before tracing. In the fourth condition vision was unrestricted.

Overshooting at a corner was found to be less when the position of the corner could be predicted either from the sequential structure of the previous part of a patterned course, or from a knowledge of the common statistical properties of the courses. Visual information acquired before tracking was found to be less effective than visual-kinaesthetic information acquired while tracking.

REFERENCES

1. POULTON, E. C. The basis of perceptual anticipation in tracking. *Brit. J. Psychol.*, 1952, **43**, 295–302.
2. SLACK, C. W. Learning in simple one-dimensional tracking. *Amer. J. Psychol.*, 1953, **66**, 33–44.
3. STEVENS, S. S. On the averaging of data. *Science*, 1955, **121**, 113–116.

(Received June 26, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

STUDIES OF DISTRIBUTED PRACTICE: XV. VERBAL CONCEPT LEARNING AS A FUNCTION OF INTRALIST INTERFERENCE

BENTON J. UNDERWOOD[1]

Northwestern University[2]

In rote learning, conditions of intralist similarity can be arranged so that the greater the intralist similarity the slower the learning (e.g. **7**). In concept learning, some evidence would also suggest that appropriate arrangement of similarity conditions would yield the same relationship (**1, 2, 4, 5**). More specifically, as similarity is increased among stimuli which are instances of different concepts, acquisition rate decreases. In the published research (cited above) on this variable, similarity has been varied, or has been inferred to vary, around a physical characteristic (e.g., size) of the stimuli per se.

In the present study, intralist similarity has been manipulated somewhat differently. An object, as symbolized by a word stimulus and representing a particular concept in the list, may have descriptive characteristics in common with other object-stimuli (representing other concepts) in the list; hence, intralist similarity can be manipulated by varying the frequency with which this commonality occurs in a given list. An illustration may make this clear. Suppose *E* arbitrarily designates three stimuli, *tomato*, *baseball*, and *collar* as instances of the concept "round." Suppose also that in the same list there are other stimuli which are used as instances of the concepts "red," "hard," and "white." The stimulus *tomato* might, therefore, be erroneously taken by *S* as an instance of "red," *baseball* of "hard," and *collar* of "white." The attainment of the appropriate groupings would be expected to be retarded in such a list as compared with one in which there was no overlap of descriptive dimensions of object-stimuli used as illustrations of different concepts. Thus, it would be predicted that the greater the number of stimuli in a list which are appropriate for more than one concept, but incorrect for all but one, the slower the learning. The present experiments test this proposition.

Evidence for facilitation of concept learning by distributed practice is weak (**5, 6**). In these two studies, however, geometric forms were used as stimuli. In order to test the generality of the conclusion concerning the influence of distributed practice on concept learning, both massed and distributed conditions have been included in the present design where verbal stimuli are employed.

Experiment I

Method

Materials.—Three lists of concepts were constructed. These lists differed only in terms of amount of intralist similarity defined by degree of overlap of descriptive characteristics for different concepts. The stimuli in the lists were nouns, the calibration for which has been published elsewhere (**10**). Essentially, for each noun the frequency with which each descriptive characteristic is used is known. These characteristics are limited to what are called sense

[1] Jack Richardson and Rudolph Schulz supervised the collection of the data and helped with the analysis.

[2] This work was done under Contract N7onr-45008, Project NR 154-057, between Northwestern University and the Office of Naval Research.

TABLE 1

LISTS USED

List 1		List 2			List 3		
Concept	Stimuli	Concept	Stimuli	Overlap R's	Concept	Stimuli	Overlap R's
Round	apple (19) derby (33) olive (17) platter (38)	Round	bean (12) belly (43) collar (16) capsule (22)	small (18) soft (24) white (44) small (51)	Round	eye (32) dime (30) pail (18) pearl (33)	small (10) small (15) small (15), white (8) small (5), white (37), hard (5)
Small	cabin (39) freckle (19) pollen (30) sardine (22)	Small	bungalow (46) kitten (25) snail (42) spot (24)	white (9) soft (41) round (14) round (21)	Small	lint (32) rice (24) tweezer (17) thimble (37)	white (38) white (54), hard (6) hard (5) round (12), hard (9)
White	diaper (50) paste (16) hospital (32) fang (10)	White	bread (35) onion (15) lard (27) salt (53)	soft (31) round (9) soft (8) small (7)	White	enamel (28) rabbit (25) skull (22) oyster (13)	hard (20) soft (18), small (6) round (11), hard (36) round (9), hard (5), small (6)
Soft	flannel (54) butter (21) custard (26) banana (11)	Soft	cradle (11) pup (12) skin (42) sheep (14)	small (24) small (50) white (9) white (23)	Hard	chestnut (18) baseball (10) stone (63) hailstone (49)	round (14), small (9) small (5), round (70), white (11) small (7), round (6) round (14), white (9), small (7)

impressions, e.g., "big," "tall," "bitter," "square," "blue," etc. For example, to the word *bean* 49% of the *Ss* in the calibration procedure responded with "green," 12% "round," 18% "small," 6% "long," and 15% of the responses were "miscellaneous," which according to the arbitrary definition includes all responses which occurred with less than 5% frequency.

In constructing the lists the objective was to keep all factors except overlap constant for the three lists. The most critical factor, insofar as can be told at the present time, is called *dominance level* of the responses to the stimuli. This is defined directly in terms of percentage frequency. Thus, as seen above, the response "green" to the stimulus *bean* has the highest dominance level of any response to that stimulus, and "long" the lowest. It has been shown elsewhere (11) that differences in mean dominance level of responses to the stimuli markedly affect the rate of learning.

The three lists as finally constructed are given in Table 1 along with supporting calibration data. Each concept has four instances in each list, and each list has four concepts, there being 16 stimuli in each list. In Column 1 under List 1, the four concepts are indicated ("round," "small," "white," "soft") with the four instances of each. The values in parentheses after each stimulus indicate the dominance level of the response to that stimulus. Thus, to the stimulus *apple*, "round" was given as a response by 19% of the *Ss*. Under List 2, the same data are given plus additional information under the column labeled "Overlap R's." To each stimulus there is a response (concept name) which is appropriate for another concept in the list. Thus, "small" was given 18% of the time to the stimulus *bean*. *Bean*, however, is used in List 2 as an instance of the concept "round." But, because four other stimuli in the list are used as instances of the concept "small" (*bungalow*, *kitten*, *snail*, *spot*) *S* may be expected to give the response "small" to stimulus *bean* as he attempts to group the stimuli under the "small" concept called correct by *E*. In List 2 each stimulus has one response which is incorrect for that stimulus but correct for four other stimuli in the list. List 1 has no such overlap; List 2 may be thought of as one degree of overlap. List 3 has still greater frequency of overlap than List 2. Some stimuli have one overlapping response, some two, and some three, for an average of 1.9.

The average dominance level of the correct responses to the stimuli in List 1 is 27.3%; for List 2 it is 27.4%, and for List 3, 28.2%. It may be noted that the concept names in Lists 1 and 2 are identical and that three of these are also used in List 3. From the materials available it was not possible to construct the three lists so that all concepts were the same in all lists.

Procedure.—A total of 306 *Ss* served in the experiment. Each list was learned by a different group of 102 *Ss*. Half of these, 51 *Ss*, learned the list with a 4-sec. intertrial interval (massed practice) and half with a 30-sec. interval (distributed practice). Three different orders

of the stimuli within each list were used with the restriction that no two instances of the same concept were presented successively. A third of the *S*s in each group started with one order of the stimuli, a third with the second order, and a third with the third order. The *S*s were assigned to a list and a condition in a rotating order based on appearance at the laboratory. Symbol cancellation was used to fill the 30-sec. intertrial interval defining distributed practice.

A trial consisted of a single presentation of all 16 stimuli. Those *S*s learning List 1 were given 10 trials, those learning List 2, 15 trials, and those learning List 3, 20 trials. These trial criteria were set in order to roughly equalize the degree of learning and were used in preference to a performance criterion in order to avoid bias due to differential loss of *S*s for inability to learn.

The method of presentation was the same as in an earlier study (11). Each stimulus was shown for 4 sec. on a memory drum. During this interval *S* responded and was immediately told "right" or "wrong" by *E*. The *S*'s response consisted of the characteristic of the object which he thought was the concept, e.g., "white." The instructions to *S* were equivalent to the "partially restricted instructions" of the earlier study. In brief, *S* was told that to the 16 stimuli there were only four correct responses, and that each response was correct for four stimuli. Illustrations of sense-impression responses were given by *S* under appropriate questioning by *E*. The *S* was told to respond each time a stimulus appeared in the window.

All *S*s returned 24 hr. following learning and were given five additional trials as retention measures.

Twenty-six *S*s were dropped from the experiments, 7 having been assigned List 1, 9 List 2, and 10 List 3. Of these, 14 were dropped because they failed to return for the retention test and the other 12 were dropped for a variety of reasons. Whenever an *S* was dropped he was replaced by the next *S* reporting for the experiment.

Results

Learning.—The two groups given List 1 had only 10 learning trials. Therefore, all six groups can be compared up through 10 trials. The first measure to be used is the mean number of total correct responses over the 10 trials. This measure simply treats the response to each stimulus item as independent from other responses. The mean totals of correct responses over the 10 trials are shown in Table 2.

TABLE 2

MEAN TOTAL CORRECT RESPONSES ON FIRST 10 LEARNING TRIALS

List	Massed		Distributed	
	M	σ_M	M	σ_M
1	71.53	4.71	73.98	4.93
2	48.52	3.52	57.94	4.38
3	57.10	3.30	57.96	3.20

Because of heterogeneity of variance, all statistical evaluations have been made by large-sample *t*'s. It will be noted in Table 2 that for all three lists performance is a little better under distributed practice than under massed. However, none of the *t*'s reached acceptable levels of significance, being .36, 1.69, and .19 for Lists 1 through 3, respectively. It must be concluded, therefore, that intertrial interval up to 30 sec. has little influence on performance on this task.

For evaluating list differences, the massed and distributed groups have been combined. Table 2 indicates that performance is superior on List 1 to that on Lists 2 and 3. The *t* between List 1 and List 2 is 4.46; between List 1 and List 3, 3.72; and between List 2 and List 3 1.23. In short, List 1 is learned significantly faster than the other two lists, but Lists 2 and 3 do not differ. These conclusions are supported by Fig. 1 which shows the mean number of items correct on each learning and relearning trial. The relearning data will be discussed later.

The above analysis does not reflect concept learning as such. That is, it tells nothing about the attainment of a concept in terms of the correct responses to all four instances of a concept. To obtain such a measure

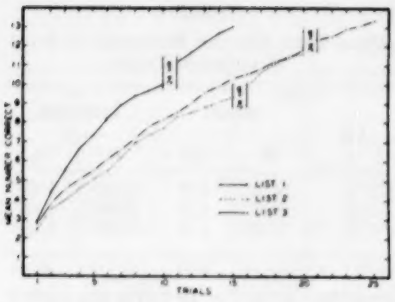

FIG. 1. Mean number of correct responses on all trials for all three lists of Exp. I.

we have taken the tenth learning trial and recorded the number of *Ss* obtaining 0, 1, 2, 3, or 4 concepts correct, where getting the concept correct is defined as giving all four correct responses on Trial 10. These data are shown in Table 3. For example, of the 102 *Ss* learning List 1, 22 got no concepts correct (all four instances) on Trial 10; 24 got one correct, and so on. A chi square for the data as given is 43.60, with 20.09 needed for the 1% confidence level. If the last two columns are combined to avoid frequencies of less than five in a cell, chi square is 39.18, with 16.81 needed for the 1% level. Clearly, more concepts were attained by *Ss* given List 1 than by those given Lists 2 and 3.

In view of the above analyses, it may be concluded that no overlap (low intralist similarity) provided an easier learning task than does a certain degree of overlap (List 2 and 3). However, since List 3 was constructed to have a greater frequency of overlap than List 2, the problem remains as to why List 3 was not more difficult to learn than List 2. In pursuit of data bearing on this issue, an error analysis was performed.

Errors during learning.—It can be seen in Table 1 that while List 3 was constructed to have greater frequency of overlapping responses than List 2, the average strength (dominance) of these overlapping responses is less than in List 2. One possibility, therefore, is that the difference in overlap did not result in greater interference in List 3 than in List 2. The basic overt-error data are presented in Table 4, where total errors over the 10 learning trials are broken down into five types as follows:

Type 1: Overlap errors scaled as having 5% or greater dominance level (as given in Table 1).
Type 2: Nonoverlap errors scaled as having 5% or greater dominance.
Type 3: Sense-impression responses not having 5% or greater dominance.
Type 4: Not sense-impression responses.
Type 5: Overlap errors not having 5% or greater dominance in scaling.

Table 4 shows that overlap errors (Type 1) occurred with greater frequency in learning List 3 than in learning List 2. Furthermore, a trial-by-trial tabulation for the first 10 trials shows that these errors are consistently greater in List 3 than List 2. It would appear, therefore, that the first objective in the list construction was realized—namely, to produce more overlap error tendencies in List 3 than List 2. Since total number of errors is only slightly greater in List 3 than List 2, it is apparent that errors of other types occur with greater frequency in List 2 than List 3. Table 4 shows that most of these errors are Type 4, errors which are not sense impressions. The most common response of this type is a superordinate response, e.g., responding with "animal" to the stimulus *puppy*. The fact that any Type 4 errors are given initially probably represents some failure in the instructions. However, it should be remembered that *Ss* were instructed to respond to every stimulus; thus many of these Type 4 errors may have resulted from the inability of *S* to think of a sense-impression response, so that rather than not respond at all, he gave some other type of response. (The percentage of "no response" varied from 10 to 20% of total

TABLE 3

NUMBER OF *Ss* ATTAINING 0, 1, 2, 3 AND 4 CONCEPTS ON TENTH LEARNING TRIAL

List	Number of Concepts Attained				
	0	1	2	3	4
1	22	24	26	15	15
2	41	40	12	8	1
3	39	39	18	4	2

TABLE 4

TOTAL OVERT ERRORS ON FIRST 10 TRIALS
(See text for explanation of types)

List	Types of Errors										Total
	1		2		3		4		5		
	f	%	*f*	%	*f*	%	*f*	%	*f*	%	
1	—	—	1558	24.4	1142	17.9	2492	39.0	1197	18.7	6389
2	2228	30.2	821	11.1	1013	13.7	2743	37.2	564	7.7	7369
3	3346	44.8	1030	13.8	918	12.3	1817	24.3	352	4.7	7463

possible responses, being low on the first trial, then increasing and finally decreasing.)

Retention.—Figure 1 showed the mean number of correct responses on each of the five relearning trials. The performance on these five trials, although occurring 24 hr. following original learning, would appear to be a simple continuation of the learning, i.e., there is no forgetting. However, a more precise analysis of the data shows that certain changes do occur over the 24-hr. period. In order to demonstrate these changes a modified form of a probability analysis has been used. In making this analysis the massed and distributed conditions have been combined. Also, Lists 2 and 3 have been combined and are compared with List 1 to determine if intralist interference, effective in learning, influences recall in any way.

The analysis consists first of determining the number of correct responses on the last trial of learning as a function of the history of each item on the preceding four trials. During these trials an item could have been given correctly (reinforced) 0, 1, 2, 3, or 4 times and the items have been so classified. Then, the percentage correct on the last learning trial for each class was determined. The resulting relationship, of course, is simply the growth in strength of an item as a function of number of reinforcements. As the second step the above procedure was repeated exactly except that the recall trial was used and the history of each item determined on the last four learning trials. The question, as far as retention is concerned, is how much if any difference is there in the two sets of data, both showing a relationship between number of reinforcements and percentage correct at recall but in one case where the "recall" was the last learning trial and in the other where it was the trial received 24-hr. later. For reasons which will not be detailed here, the 24-hr. recall will be slightly overestimated relative to the last-trial recall. However, the main concern is comparison between List 1 and combined Lists 2 and 3.

The percentage correct and the percentage of errors after the varying number of reinforcements are shown in Fig. 2. It will be noted first in Fig. 2 that the only case in which there is an actual increase in correct responses over 24 hr. as compared with the few seconds between the last

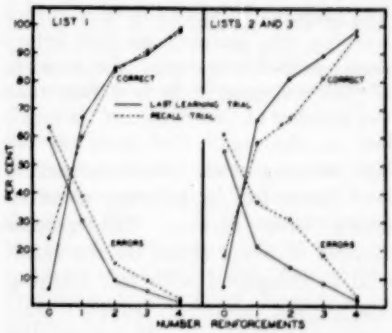

FIG. 2. Correct responses and overt errors on last learning trial and on recall trial as a function of number of reinforcements on previous four trials for List 1 and for Lists 2 and 3 combined.

and next to the last trial is the one in which the item had never been reinforced. This is true for List 1 and for Lists 2 and 3. Second, it should be noted that the percentage correct recall on last learning trial as a function of number of reinforcements is almost exactly the same for items of List 1 and for those of Lists 2 and 3. Third, since these growth functions are so nearly identical, loss over 24 hr. for Lists 2 and 3 may be taken as indication of forgetting. Furthermore, since no such loss is apparent for List 1 it may be concluded that differences in intralist similarity are responsible for the differences in retention.

It was noted that items which had never been reinforced showed a greater gain over 24 hr. than over the few seconds between the last two trials of learning. There are two factors believed to lie behind this fact. First, *Ss* rehearse; they report that they do by way of "thinking about" items which they had not gotten correctly during the learning session. Secondly, *Ss* are instructed just before recall to be sure and give a response to all stimuli. While they had been instructed to do this during learning, the instructions just before recall probably reduced the number of "no response." It is a fact that the number of "no response" is much less on the recall trial than on the last learning trial. This instruction may have led to guesses, some of which were correct. The greater number of overt errors on the recall trial as compared with last learning trial would seem to support this guessing hypothesis. However, the main point to be made by Fig. 2 is that there is some forgetting of these materials when the intralist interference is high.

Discussion

The data have shown that learning rate of these verbal concepts is not influenced by distributed practice. Actually for all three lists distributed practice was slightly superior to massed although not significantly so. The results thus approximate those found with other concept-formation tasks (**5, 6**). Furthermore, they parallel those usually found with rote paired-associate lists (e.g. **8**). When distributed practice *does* facilitate learning of paired-associate lists it is when rate of presentation is very rapid (**3**) or when material is of very low meaningfulness (unpublished). In view of the fact that there is some operational comparability between paired-associate learning and concept learning, any subsequent studies on the massed-distributed variable might well increase rate of presentation, reduce meaningfulness, or both. However, it should be recognized that by so doing more and more emphasis may be placed on the rote component of the task. In any event, on the basis of the data now available, it appears that when distributed practice is defined in terms of relatively short intervals between trials it is not a relevant variable in concept learning.

List 1, with low intralist interference, was acquired more rapidly than Lists 2 and 3 in which intralist interference was higher. It was expected that List 3 would be more slowly learned than List 2 because of the greater overlap among response tendencies. This difference in learning was not found, although the greater overlap for List 3 was reflected in the greater error frequency for the overlap response tendencies in this list. Now, it is quite possible that increasing error tendencies beyond a given point will not appreciably affect rate of learning; this has been found in some paired-associate studies (e.g. **8, 9**). However, since the present concept-formation study uses relatively new material there is a distinct possibility that the lists might have differed on some characteristic in addition to degree of overlap.

Such a danger is suggested by the difference in number of Type 4 errors for Lists 2 and 3. These Type 4 errors are responses that are not sense impressions; List 2 resulted in more of these than did List 3. It is therefore possible that the stimuli of List 3 have more "power" to elicit these sense-impression responses than do those of List 2. Since the sense-impression responses are the class called correct in the present study it is possible that the greater overlap of concept characteristics in List 3 was counteracted by the fact that List 2 had less tendency to elicit sense-impression responses. While the error data suggest this possibility, it must be clearly recognized that when *S*s are instructed to respond to every stimulus (even though they don't) there will inevitably be a certain amount of reciprocity among the types of errors. Thus, if List 3, due to the largest overlap, elicits a large number of overlap errors, it must inevitably follow that there will be fewer of the other types, perhaps of Type 4. That such an argument has some plausibility is seen by the fact that List 1 actually produced the greatest frequency of Type 4 errors. And, it is possible that the overlap tendencies of List 3 might be expected to result in fewer Type 4 errors than the other lists because these overlap error tendencies would restrict the range of *S*s responses.

It appears that from the present data no decision can be reached as to whether or not the results might be attributed to some characteristic other than the manipulated variable. The intent of the research was to study rate of concept formation as a function of degree of overlap (intralist interference). The above considerations leave some doubt as to whether or not the lists were equal on all characteristics other than overlap. To remove such doubt insofar as it is possible in terms of our present understanding of the material used, another study has been performed. In this experiment *S* was given the four correct responses before learning the list. It was reasoned that by so doing differences among lists other than overlap tendencies would be minimized. Thus, if lists differed in their tendency to elicit sense impressions, giving the *S*s the responses would eliminate the affect of such differences on learning.

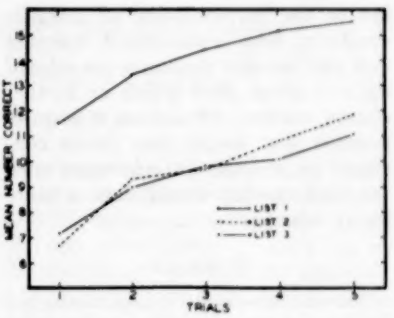

FIG. 3. Mean number of correct responses on the five learning trials of Exp. II for all three lists.

EXPERIMENT II

Method

The same three lists were used as in Exp. I. Before *S* started learning the list he was presented a card on which the four correct responses appeared. The *S* was told that these were the only four responses with which he was to be concerned. He studied these four responses until he could recite them to *E*. Furthermore, the card was left in front of *S* during the learning period. The instructions otherwise were the same as for the first experiment. A total of 72 *S*s were used, 24 for each list. The *S*s were assigned in rotation to a list in order of appearance at the laboratory. Five acquisition trials were given but no retention measures were taken.

Results

The mean number of correct responses on each trial for each list is shown in Fig. 3. It is apparent that the results correspond almost exactly to those found in the first experiment. Performance on Lists 2 and 3 is inferior to that of List 1, but performance on Lists 2 and 3 does not differ. These results make it seem unlikely that the lists differ on a relevant characteristic other than degree of overlap. The conclusion is that as

far as the manipulation of intralist similarity here is concerned, learning rate and intralist similarity are related up to a point after which no further change occurs. Of course, it is quite possible that longer lists (more concepts) or stronger interferences with the high overlap would show a more linear relationship.

Summary

This study investigated concept learning as a function of degree of intralist interference and intertrial interval. Verbal concept materials were employed in which intralist interference was varied in terms of number of common characteristics shared by the objects symbolized by their respective word stimuli when those stimuli were instances of different concepts. Three lists were constructed varying in degree of such overlap among concept stimuli. Massed practice consisted of 4 sec. between trials, distributed practice involved 30 sec. between trials. Six groups of *Ss*, 51 in each group, were used. The six groups represent the six conditions, namely, three degrees of intralist similarity with a massed and distributed group for each degree. The *Ss* were given a constant number of learning trials with retention being measured after 24 hr.

The major results were:

1. The greater the overlap (intralist interference) the slower the learning up to a point. Beyond this point with the present materials no further change occurred. A check experiment in which *Ss* were given the correct responses before learning makes it seem unlikely that the lists were different on any critical characteristic other than degree of overlap.

2. Confirming previous concept-formation studies, and studies using rote-learned paired associates, no differences in learning occurred as a consequence of massed vs. distributed practice.

3. The forgetting over 24 hr. was greater for lists with high intralist similarity than for the list with low intralist similarity, but was small in all lists. Absolute losses were difficult to determine because of rehearsal and guessing problems.

References

1. Baum, M. H. Simple concept learning as a function of intralist generalization. *J. exp. Psychol.*, 1954, **47**, 89–94.
2. Buss, A. H. A study of concept formation as a function of reinforcement and stimulus generalization. *J. exp. Psychol.*, 1950, **40**, 494–503.
3. Hovland, C. I. Experimental studies in rote-learning theory: VIII. Distributed practice of paired associates with varying rates of presentation. *J. exp. Psychol.*, 1949, **39**, 714–718.
4. Newman, S. E. The effects of similarity and contiguity on the formation of concepts. Unpublished doctor's dissertation, Northwestern Univer., 1951.
5. Oseas, L., & Underwood, B. J. Studies of distributed practice: V. Learning and retention of concepts. *J. exp. Psychol.*, 1952, **43**, 143–148.
6. Richardson, J., & Bergum, B. O. Distributed practice and rote learning in concept formation. *J. exp. Psychol.*, 1954, **47**, 442–446.
7. Underwood, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44**, 80–87.
8. Underwood, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–142.
9. Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149.
10. Underwood, B. J., & Richardson, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84–95.
11. Underwood, B. J., & Richardson, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, **51**, 229–238.

(Received June 28, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

DIFFERENTIAL SENSITIVITY IN GUSTATION [1,2]

HOWARD G. SCHUTZ [3] AND FRANCIS J. PILGRIM

Quartermaster Food and Container Institute for the Armed Forces

In the 88 years that published reports on gustatory differential sensitivity have appeared, only two to our knowledge cover studies on the four accepted primary taste qualities. Keppler (**10**) in 1869 tested the four, but he himself was his only *S*. Trout and Sharp (**21**) reported work in 1937 on all four qualities using four *S*s, but they used ranking and correlation to obtain their measures of sensitivity so that one cannot calculate a difference limen (ΔI) from their data. The literature (**2, 7, 9, 10, 12, 13, 17**) on the Weber ratio ($\Delta I/I$) for taste shows wide variation in the values obtained—the $\Delta I/I$ for each quality differing about tenfold. In addition, the question of the constancy of the Weber ratio has been raised repeatedly, with contradictory answers (**8, 16**).

The present study was designed to determine for the four basic taste qualities at five levels of intensity (I): (*a*) differential sensitivity, $\Delta I/I$; (*b*) the relationship of $\Delta I/I$ to I; (*c*) the relative order of the sensitivities; and (*d*) individual differences in $\Delta I/I$.

METHOD

Psychophysical method.—Pfaffmann (**14**) demonstrated that the method of constant stimuli and the method of single stimuli yield almost identical results. He also concluded that since the method of single stimuli is more convenient than the method of constant stimuli it can be used to great advantage in taste research. Pfaffmann used six stimulus points. In order to determine whether or not fewer points would give comparable results, a preliminary study was conducted in which the method of single stimuli was used to determine the ΔI for .95% sodium chloride. One group of 10 *S*s judged six points and another group of 10, four points. The results were almost identical: ΔI for four points, .10, for six points, .12. Therefore, four stimulation points were considered sufficient.

Next, the concentration interval between the stimulus solutions had to be selected. In order to insure equality of judgment throughout the experiment, one interval was used for all qualities and all levels of I. This procedure also allows combining the data across qualities or levels. The interval selected was .15; in terms of a (hypothetical) concentration midpoint (CM) of 1.00, the stimuli, designated 1, 2, 3 and 4, were thus, .775, .925, 1.075, and 1.225 of the CM. The .15 interval is small enough to maintain the attention of *S*, and large enough to cover the anticipated range of sensitivity.

Application of stimulus.—An integrating technique similar to that which Cameron (**3**) recommends was utilized since it appeared to have the least number of objections (**19**) and is also most like normal tasting. Six milliliters of the test solutions were delivered from a calibrated automatic pipette into one ounce "shot" glasses which had been washed and rinsed in distilled water and dried in an oven for 20 min. A previous study (**18**) indicated that 6 ml. was an acceptable amount of solution for *S*s to hold in their mouths (i.e., did not give a

[1] This paper reports research undertaken at the Quartermaster Food and Container Institute for the Armed Forces, and has been assigned Number 577 in the series of papers approved for publication. The views or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the views or indorsement of the Department of Defense.

The authors are indebted to Kenneth R. Wood of the Statistics Office for his valuable advice and assistance.

[2] Part of this material was presented at the 63rd Annual Meeting of the American Psychological Association, San Francisco, Calif., Sept. 7, 1955.

[3] The research reported in this paper is based on a dissertation submitted by the senior author to the Graduate School of Illinois Institute of Technology in partial fulfillment of the requirements for the Ph.D. degree. He is indebted to his adviser Dr. Marion H. Groves for his suggestions and encouragement throughout the research.

feeling of "overcrowding") and was sufficient to thoroughly stimulate the buccal cavity.

Adaptation varies with the method of application of the stimulus. The *Ss* were instructed to swirl the test solution around in the mouth and to expectorate and rinse afterwards so as to minimize adaptation. The *Ss* differed from one another in their stimulation time from 5 to 10 sec. The period between stimulations was 30 sec. An *S* tested 20 solutions in approximately 15 min., depending on his speed of operation.

There are various techniques and conflicting opinions (**1, 2, 3, 4, 14, 20**) on the most desirable method of rinsing. The *Ss* were instructed to rinse their mouths with water in a consistent and natural way after each stimulation, but the amount and number of times were left to their own judgment. Rinsing time varied from about 5 to 15 sec.

Levels of intensity.—Two kinds of preliminary experiments were conducted in order to establish the lower and upper limits of *I* to be tested. These limits were set as the absolute thresholds and the stimuli that were subjectively very strong. Since the lowest and highest levels of *I* are represented by the CM, the actual limits are Stimulus 1, or .775 of the lowest CM, and Stimulus 4, or 1.225 of the highest CM.

The absolute thresholds for the four substances were obtained by the method of constant stimuli with 10 *Ss*. The results were as follows: sodium chloride, .089%; citric acid, .004%; sucrose, .35%; and caffeine, .022%. To determine the highest stimuli that should be used, 10 *Ss* gave their free comments on three increasingly strong solutions of the four taste substances. The solutions were tested in random order. The value for the strongest stimuli to be tested was chosen on the basis of the comments elicited. The two extremes and the three intermediate levels for each quality were then spaced approximately logarithmically. The midpoint values are shown in Table 1.

The final range of concentrations selected for each of the qualities compares favorably with the most extensive of previous experiments in this area, considering the fact that 6 ml. of solution was used to stimulate rather than one drop as in the work of Holway and Hurvich (**9**). The range is wide enough to both cover the intensities that people experience in everyday life and also adequately test Weber's law.

Materials.—Water for mouth rinsing and preparation of solutions was distilled and then percolated through activated charcoal to remove taste and odor.

C.P. sucrose, sodium chloride, and citric acid, and U.S.P. caffeine were used, each drawn from a single lot for the entire experiment. These compounds are commonly encountered in ordinary eating and also elicit "pure" taste qualities. The materials were weighed to three significant digits and made up to final volume (% wt./vol.). Several times during the experiment the pH of the water was measured and found to remain within the limits 6.0±.3. This small difference would not affect the results.

Subjects.—Ten *Ss*, five male and five female, were utilized throughout the experiment on differential sensitivity. They were all volunteers, and were employees of this Institute. They ranged in age from 20 to 58 yr. with a mean age of 33 yr. The *Ss* were all practiced in making psychophysical judgments of the type required.

Controls.—All testing was conducted between 3:00 and 4:00 P.M. The *Ss* sat in individual booths in an air-conditioned room at 24° C. ± 2°C. and a relative humidity of 35–50%. The *Es* were located on the other side of a partition with a turntable in it through which *Es* presented samples and *Ss* transmitted their responses. The samples were of 6 ml. volume. All solutions, as well as the rinse water, were served at room temperature where differences in temperature have little effect on sensitivity (**11, 21**). The *Ss* were instructed to refrain from eating, drinking (except water), or smoking for one-half hour before testing.

The *Ss* did not know the order in which the qualities and levels of intensity were to be tested, thus minimizing expectation effects. As a check on the possible effects of time, practice, or fatigue, a retest of the first differential threshold, without the knowledge of the subjects that it was a repeat, was run as the last threshold determination.

Procedure.—Each threshold determination consisted of two sessions on consecutive days. At each session the 10 *Ss* each made judgments on four sets of test solutions. Each set consisted of the four equally spaced concentrations randomly presented, making a total of 80 responses per stimulus for each level of intensity of each quality. The 20 threshold determinations (five levels for each of four qualities) were tested in a random order with the restriction that one quality could not be tested more than two consecutive times.

All experimental sessions were completed in a period of three months. Missed sessions were made up in their proper sequence.

Next to each turntable was a rack holding four stacks of brass tags numbered 1, 2, 3, or 4. There were always at least six of each number on every rack. Also, in each booth there was a glass of rinse water which *S* could refill.

At the beginning of a session, *S* was informed which quality was being tested. He first received the four solutions one at a time (of the

set comprising that day's level of intensity and quality) in ascending order, from 1, the weakest, to 4, the strongest. He was told that this set of solutions was a practice set to acquaint him with the range of solutions and their relative strength. The *S* then received four sets identical to the practice set but with the stimuli in a random order. This randomization was carried out for each *S* on every set of four stimuli with the restriction that no *S* was presented stimuli in the same order for any one of the eight sets of unknowns for each threshold. The *S* assigned each unknown stimulus a number from 1 to 4 to indicate his subjective estimate of its relative intensity. A number could be used more than once in any given set, but no corrections were allowed; thus the probability of choosing any one number was equal for every solution in the set.

The *S* tasted each solution by swirling all of it around in his mouth; then he expectorated it into a dental cuspidor and rinsed his mouth. He placed the empty "shot" glass on the turntable along with a brass tag (except for the practice set) and turned the turntable toward an *E*. At this time the *E* started timing the interval between presentations. At the end of 30 sec. the next sample was presented. Between each set of four stimuli there was also a timed interval of 30 sec., during which *S* replaced the tags on the rack.

Treatment of data.—The raw data in this experiment consisted of judgments of 1 to 4 for each stimulus.[4] A judgment of 3 or 4 was scored as a plus. In this way, the number of times out of 8 that a stimulus was called greater than the CM was determined for each of the four stimuli for a given quality, level, and *S*.

The combining of judgments is possible because the concentration interval for all qualities and levels of intensity is a constant ratio of .15. In this manner an average difference threshold could be computed for any combination of qualities, levels, or people.

For combined judgments the proportions were converted to normal deviates. In two cases of zero proportion the deviates were estimated. For individual data with 8 judgments per point the binomial expansion was computed to obtain the equivalent of the normal deviate with allowance for proportions of 0 and 1. Fitting a line to the deviates by the method of least squares yielded the slope of the line, b, and the variance of the slope, s^2_b (**22**). The equation of the line can then be solved to yield the point of subjective equality (PSE), ΔI and $\Delta I/I$. By obtaining the ΔI from the PSE rather than from the CM, the ΔI is an average of the lower and upper ΔI's.

The slope, as a measure of discrimination, is inversely proportional to $\Delta I/I$. The use of b is preferable because it has an empirical error, s^2_b, whereas $\Delta I/I$ does not; and this error is uncorrelated with b, whereas the standard error of a proportion, $\Delta I/I$, is a function of the size of the proportion. One disadvantage is that the PSE is not accounted for by b; however, this presents no problem with the present data because of the insignificantly small variations of the PSE from the CM, as is shown later.

The use of b presents one problem that requires treatment of the data to take two forms. The average b based on b's obtained from small numbers of judgments differs from the b obtained after cumulating the judgments as is ordinarily done in psychophysical studies. The average b is generally the larger and leads to the conclusion that the sensitivity is greater than it actually is. However, the 200 individual b's for each quality, level, and *S* are required for a complete analysis of variance. Therefore, the psychometric values are represented by b's obtained from judgments cumulated over subject, quality, or level. The analysis of variance is based on individual b's. It was demonstrated empirically that analyses of variance of b's based on cumulated judgments and with the error term based on the average s^2_b of the b's in the analysis yield F ratios of the same order of magnitude and significance at approximately the same level as those presented (Table 3). Therefore, it is felt that inferences drawn from the analysis may be applied to the psychometric values presented in traditional form with little risk of error.

Results

The results are reported in terms of the following psychometric values: PSE, ΔI, $\Delta I/I$, and b. The results for the group are given in Table 1. The sequence in which each threshold was determined can also be found in this table. Two general facts are apparent: sensitivity to bitter is poorest at all levels and sensitivity to

[4] The raw data have been deposited with the American Documentation Institute. Order Document No. 5271 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

PSYCHOMETRIC VALUES FOR THE GROUP OF 10 SUBJECTS FOR EACH QUALITY AND LEVEL

Quality	Level	Test Sequence	Concentration Midpoint (%)	PSE	ΔI	$\Delta I/I$	b
Salt	1	7	.150	.150	.0289	.193	.525
	2	10	.400	.396	.0518	.131	.780
	3	2	1.10	1.10	.133	.121	.837
	4	13	3.00	2.98	.357	.120	.850
	5	20	8.00	7.80	1.81	.231	.448
Sweet	1	17	.500	.496	.150	.303	.337
	2	11	1.50	1.53	.221	.144	.687
	3	14	3.00	3.07	.481	.157	.631
	4	8	8.00	7.97	1.23	.154	.658
	5	3	20.0	19.8	2.87	.145	.705
Sour	1	12	.010	.0100	.0029	.290	.350
	2	18	.030	.0303	.0069	.228	.443
	3	1	.100	.101	.0235	.232	.430
	4	6	.300	.300	.0510	.170	.596
	5	16	1.00	1.01	.227	.224	.446
Bitter	1	5	.0315	.0317	.0122	.385	.259
	2	19	.0600	.0607	.0206	.339	.295
	3	15	.125	.125	.0332	.265	.381
	4	4	.250	.248	.0715	.288	.354
	5	9	.500	.503	.124	.248	.406

all qualities is best in the middle range of intensities.

The error of the PSE from the CM (assigned a value of 1.0) was found for each quality and level. The 20 PSE's had a mean of 1.002 and a range from .975 to 1.023 which represents an *SD* of only .5% from the midpoint.

Combinations of the data over levels yielded 400 judgments per stimulus point for each quality and 1600 judgments per stimulus point over all qualities. In Table 2 are the psychometric values for each of the four taste qualities based on data from all levels of intensity and the psychometric values over all qualities and intensities, we might say for gustation as a whole.

Table 3 gives the results of the analysis of variance. The fitting of curves by use of orthogonal polynomials (6) helps reveal where the variation occurs across levels within each quality. Only linear and quadratic effects were considered. Bitter does not differ significantly by level; i.e., the Weber ratio is constant for this quality. For sweet and sour there is a linear slope differing from the horizontal; however, the difference for sweet is due only to poor sensitivity near absolute threshold—the second through fifth levels have essentially the same $\Delta I/I$ or b. In the case of salt, a quadratic curve explains most of the variation. Here again, there is a loss of sensitivity near absolute threshold, but there is also a loss at the highest intensity where the *S*s reported feeling mainly a stinging sensation, in addition to the salty taste. The highest level of bitter and sweet did not give qualitative modulations. The highest level of sour produced some feeling, but this did not significantly affect sensitivity as is evidenced by the lack of a quadratic effect.

In order to locate the between-quality significances found by analysis

TABLE 2

PSYCHOMETRIC VALUES FOR EACH QUALITY OVER ALL LEVELS OF INTENSITY AND FOR ALL QUALITIES

Quality†	Proportion Judged Stronger than CM*				PSE‡	ΔI	$\Delta I/I$	b	s_b
	1	2	3	4					
Salt	.142	.442	.638	.840	.993	.152	.153	.668	.0036
Sweet	.192	.378	.565	.832	1.008	.173	.172	.586	.0017
Sour	.240	.388	.622	.722	1.006	.226	.224	.449	.0018
Bitter	.302	.418	.592	.678	1.003	.303	.302	.338	.0008
All Qualities§	.219	.406	.604	.766	1.002	.202	.202	.500	.00014

* 1, 2, 3, 4, refer to concentrations 1, 2, 3, 4 at all levels. Interval was .15 in all cases.
† Based on 1600 judgments, or 400 per stimulus point.
‡ Computed with 1.00 as CM.
§ Based on 6400 judgments.

of variance, a Duncan multiple range test (5) was applied. Bitter (.454) differed from sweet (.856) and salt (1.047) at the .01 level; sour (.611) differed from sweet (.856) at the .05 level and from salt (1.047) at the .01 level; bitter did not differ significantly from sour, nor did sweet differ significantly from salt.

The psychometric values for each subject for the four qualities are shown in Fig. 1.

DISCUSSION

Comparison with previous results.—In terms of order of sensitivity to qualities, from most to least sensitive, the present results differ from the order that Bujas (2) reports—sour, salt, and sweet; and from Kopera (12)—using Saidullah's value for salt—who gives the order as sweet, salt, and bitter. The position of bitter as the least sensitive is the only result that remains constant from experiment to experiment, except for the work of Keppler (10) in which bitter is found to be the most sensitive. In considering the work of Keppler it must be remembered that he used only one *S*—himself.

Of the previous experiments Saidullah (17), Kopera (12), and Bujas (2) found that $\Delta I/I$ was constant over nearly the entire range that was tested, while Lemberger (13), Holway and Hurvich (9), and Fodor and Happisch (7) reported that it was constant only in the middle range. Keppler (10) did not test over a wide enough range to come to any conclusion about Weber's law. Only Bujas, Holway and Hurvich, and Lemberger tested in the area of the absolute threshold.

Another comparison with previous work that can be made is of the actual values reported for $\Delta I/I$ regardless of the levels of intensity that were tested. The present ratio for salt, .15, agrees most closely with the work of Saidullah,

TABLE 3

ANALYSIS OF VARIANCE OF SLOPES

Source	*df*	*MS*	*F*	*P*
Quality*	3	3.436	15.27	.001
S†	9	1.205	8.55	.001
S-Quality†	27	.225	1.60	.05
S-Level†	36	.243	1.72	
Level within Quality†				
Salt–linear	1	.002	<1	
Salt–quadratic	1	5.091	36.10	.001
Sweet–linear	1	1.254	8.89	.01
Sweet–quadratic	1	.905	6.42	.05
Sour–linear	1	.653	4.63	.05
Sour–quadratic	1	.056	<1	
Bitter–linear	1	.379	2.69	
Bitter–quadratic	1	.000	<1	
Residual‡	8	.187		
S-Quality-Level	108	.141		
Total	199			

* Error term is S × Q interaction.
† Error term is S × Q × L interaction.
‡ Contains cubic and quartic variance for all qualities.

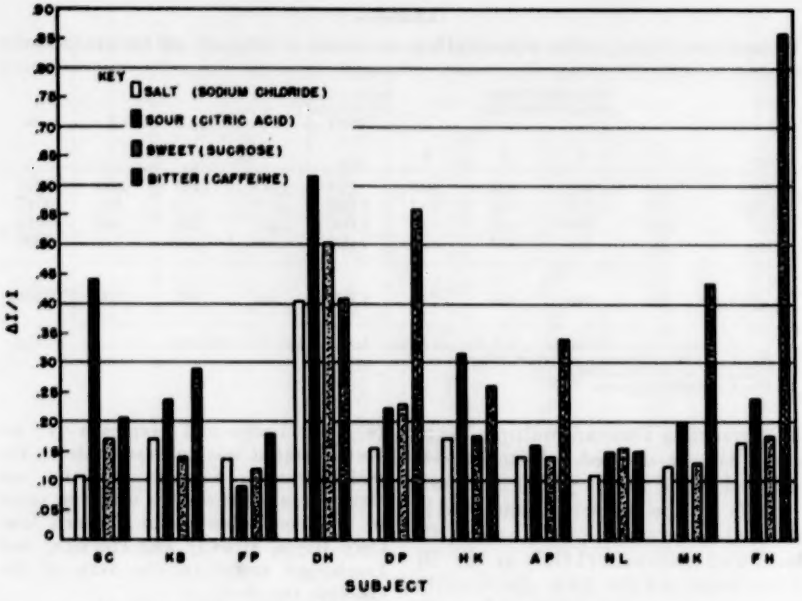

FIG. 1. Differential sensitivity for the four basic taste qualities over all levels of intensity for each of 10 *Ss*.

.14, and Fodor and Happisch, .15, and least with the work of Keppler, .67, Bujas, .25, and Holway and Hurvich, .50 (calculated average), with a minimum value of .24. In all these experiments the substance used was sodium chloride, although the purity of the chemical is not stated by any but Holway and Hurvish.

Sour can be compared only with the work of Bujas, for in the only other experiment on sour a different acid was used (Keppler). Bujas' result of .20 is not far from the result of .24 obtained in the present experiment. Keppler found a Weber ratio of .04 for phosphoric acid, a sensitivity that is not at all in line with the present results or those of Bujas.

For the sweet quality the present study yields a Weber ratio of .17. This figure is in close agreement with the results of Lemberger, .20, and with Krogh-Jensen's value of .21, as quoted in (12) and (17), but not at all with the results of Keppler, who used glycerine rather than sucrose, .04, and only roughly with Kopera, .11, or Bujas, .32.

Bitter can be compared with the results of only two *Es*, Keppler and Kopera. In both these experiments quinine was the stimulus for the bitter taste, so comparison with the present study in which caffeine was the bitter stimulus may not be valid. The present ratio of .30 is higher than the .21 found by Kopera and much higher than Keppler's value, .02. The purity of the chemicals was not stated by these two *Es*.

There are several possible explanations for the differences in results between the present experiment and the previous experiments in this area. The first is in the area of psychophysical methodology because the methods differ widely. A second is the range of intensity tested; the present study revealed differences with levels for some qualities, especially at the low and high intensities. A third reason may be the differences in the

purity of the chemicals used, or in the case of bitter, the difference between caffeine and quinine. Lastly, the number of *S*s varied from experiment to experiment, a factor which may be of great importance, as is indicated by the individual differences discussed below.

Possible sources of errors.—Practice effects and motivation were studied by repeating the first determination, the third level for sour, at the end of the series, which was three months later, without *S*s knowing that it was a repeat test. The b's and s^2_b's were .430±.069 and .421±.078 and the PSE's, 1.01 and .99. These results have been taken as an indication of the stability of performance throughout the test.

Other factors that could conceivably influence the results are subjective reactions to the solutions either in terms of preference, mainly dislike for some stimuli, especially the strong ones, and also the subjective intensity per se. Therefore, after the main part of the study, *S*s rated, during two sessions, the midpoint concentrations of the 20 sets of stimuli on a 9-interval intensity scale. At another two sessions they rated the same samples on a 9-interval preference scale. In each case, the 20 samples were randomized. There were no significant correlations between group sensitivity and either preference ($r = .33$) or subjective intensity ($r = .11$).

There is reasonable certainty, then, that the results are independent of affective reactions to the stimuli, of practice and motivation effects, and of time errors in the method. Postexperimental checks also indicated that none of the stimulus solutions had any odor, so that no errors could have arisen from this source. Although data are limited, there was no evidence of any effect due to age, pH of saliva, or smoking habits.

Individual differences.—In previous experiments the number of *S*s varied from one to eight. As far as could be ascertained there has not been a study in which individual differences in differential sensitivity to the gustatory sense have been determined. In the present experiment it was possible to obtain a measure of the individual *S*'s differential sensitivity by combining his judgments over all levels of intensity for each of the four qualities. The results are given in Fig. 1. Each quality has a fairly normal distribution of b values. Bitter has the smallest variation (s^2), then in order come sweet, salt, and sour. From inspection of Table 3, it becomes apparent that there are definite individual differences in differential sensitivity. The analysis of variance shows a significant F for S as a source of variation. This analysis also revealed, by the significant F for the S-quality interaction, that not all *S*s were sensitive to the four taste qualities in the same order. This latter result contributes to the idea of the uniqueness of the four taste qualities. However, an analysis that indicates the extent of the relationship among the qualities are the six correlations, each with an N of 10, between *S*s' b values for the four qualities. The only two quality comparisons to approach a significant correlation were between sweet and sour, $r = .65$. However, they are all positive, the smallest r being .37.

The results on individual differences showed that the highest S variability to the four qualities was for sour. This may be interpreted on the basis of the studies on specific fibers (**15**) which showed that only acid stimulated three types of fibers. It is possible that the greater number of fibers involved in the response to acid, as compared with the other substances, leads to the greater probability of individual differences in sensitivity. It is recognized, of course, that the specific fiber work was done on a nonhuman species and that the human fiber organization for taste may be quite different.

Summary

Differential sensitivity to the four primary taste qualities at five levels of intensity was determined by the method of single stimuli on 10 *S*s. Differences among the differential sensitivities to the four taste qualities are significant. Over all levels, $\Delta I/I$ for salt is .15; for sweet, .17; for sour, .25; for bitter, .30; and for gustation as a whole, .20. Differential sensitivity for

the levels of intensity tested, within each quality, does not differ for bitter and only differs near the absolute threshold for salt, sour, and sweet and near the "terminal" threshold for salt. Individuals differ widely in their differential sensitivity to the basic tastes within the qualities and in the order to which they are sensitive to different qualities.

References

1. Biester, A., Wood, M. W., & Whalin, C. S. Carbohydrate studies. I. The relative sweetness of pure sugars. *Amer. J. Physiol.*, 1925, **73,** 387–396.
2. Bujas, Z. La mesure de la sensibilitié differentielle dans le domaine gustatif. *Acta Inst. Psychol.*, 1937, **2,** 1–18.
3. Cameron, A. T. The taste sense and the relative sweetness of sugars and other sweet substances. *Scient. Rep. Sugar Res. Found.*, New York, 1947, No. 9.
4. Dahlberg, A. C., & Penczek, E. S. The relative sweetness of sugars as affected by concentration. *N. Y. St. Agric. Exp. Sta. tech. Bull.*, 1941, No. 258.
5. Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11,** 1–42.
6. Fisher, R. H., & Yates, F. Y. *Statistical tables.* New York: Hafner, 1953.
7. Fodor, K., & Happisch, L. Über die Verschiedenheit der Unterschiedsschwellen für den Geschmackssinn bei Reizzunahme und Reizabnahme. *Pflüg. Arch. ges. Physiol.*, 1922, **197,** 337–347.
8. Holway, A. H., & Pratt, C. C. The Weber-ratio for intensitive discrimination. *Psychol. Rev.*, 1936, **43,** 322–340.
9. Holway, A. H., & Hurvich, L. Differential gustatory sensitivity to salt. *Amer. J. Psychol.*, 1937, **49,** 37–48.
10. Keppler, F. Das Unterscheidungsvermögen des Geschmacksinn für Konzentrationsdifferenzen der Schmeckbaren Körper. *Pflüg. Arch. ges. Physiol.*, 1869, **2,** 447–458.
11. Komuro, K. Le sens du goût a-t-il un coefficient de temperature. *Arch. Neer. de Physiol. de l'Homme et des Animaux*, 1921, **5,** 572–579.
12. Kopera, A. Untersuchungen über die Unterschiedsempfindlichkeit im Bereiche des Geschmacksinns. *Arch. ges. Psychol.*, 1931, **82,** 272–307.
13. Lemberger, F. Psychophysische Untersuchungen über den Geschmack von Zucker und Saccharin. *Arch. ges. Physiol.*, 1908, **123,** 293–311.
14. Pfaffmann, C. An experimental comparison of the method of single stimuli and the method of constant stimuli in gustation. *Amer. J. Psychol.*, 1935, **37,** 470–475.
15. Pfaffmann, C. Gustatory afferent impulses. *J. cell. comp. Physiol.*, 1941, **17,** 243–258.
16. Pfaffmann, C. Taste and smell. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1145–1172.
17. Saidulbah, A. Experimentelle Untersuchungen über den Geschmacksinn. *Arch. ges. Psychol.*, 1927, **60,** 457–484.
18. Schutz, H. G. The influence of an olfactory stimulus upon a gustatory threshold. Unpublished master's dissertation, Illinois Inst. of Technology, 1952.
19. Scofield, E. S. An experimental study of some factors influencing the determination of gustatory thresholds. Unpublished doctor's dissertation, Ohio State Univer., 1939.
20. Skramlik, E. v. *Handbuch der Physiologie der niederen Sinne:* Bd. 1, *Die Physiologie des Geruchs-und Geschmackssinnes.* Leipzig: Thieme, 1926.
21. Trout, C. M., & Sharp, P. F. The reliability of flavor judgments, with special reference to the oxidized flavor of milk. *Cornell Univer. Agric. Exp. Sta.*, 1937, No. 204.
22. Youden, W. J. *Statistical methods for chemists.* New York: Wiley, 1951.

(Received June 29, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

EFFECT OF REDUCED EXPOSURE DURATION ON BRIGHTNESS CONSTANCY [1]

H. LEIBOWITZ AND P. CHINETTI

University of Wisconsin

Experimental manipulation of exposure duration has been of value in the systematization and interpretation of the processes involved in the end-organ activity of the visual system (**3, 4**). Similarly, for the perception of shape, size, and brightness, variation of exposure duration has proved valuable by demonstrating a differential effect depending upon the visual measure under consideration. Reduction of exposure duration reduces the tendency toward shape constancy, has little effect on size constancy, and increases the tendency toward brightness constancy (**8, 9**). The present study represents a detailed investigation of the effect of the time variable on brightness constancy.

In a previous experiment (**7**), the role of simultaneous brightness contrast on brightness constancy was investigated by matching a test object, viewed over an illuminance range of more than 1,000,000 to 1, with a photometric field under different conditions of test-object contrast. Contrast was defined in terms of several recent experiments (**1, 2, 5, 6**). These studies indicate that when two stimuli of different luminance values co-exist in the visual field, the field of lower luminance appears less bright while the change in the appearance of the field of higher luminance is negligible. Simultaneous contrast, therefore, is seen to result primarily from the inhibition of the field of lower luminance by the field of higher luminance. The role of simultaneous contrast, as here defined, in brightness constancy has been demonstrated by the observation that a gray test-object viewed against a less or equally luminant surround exhibits some tendency toward brightness constancy. However, if the test object is viewed against surround of higher reflectance, thereby introducing test-object contrast, the tendency toward constancy is increased and the matches are more in agreement with the "law" of brightness constancy (**7**).

In the present experiment, matches are obtained for both a continuous and an extremely brief exposure duration of the test object with simultaneous contrast as parameter. Such data permit analysis of the differential effect of reduced exposure duration on the constancies as well as the mechanism underlying the importance of simultaneous contrast in brightness constancy.

Apparatus and Procedure

The apparatus has previously been described in detail (**7**). It consists of two adjacent, equal-sized chambers, one containing the test object and the other a photometric field. The test object, mounted at eye level on the wall of the stimulus chamber, consists of a square Munsell gray paper subtending a visual angle of 1° on a side at the viewing distance of 6 ft. The *S*, by means of a slight head movement, may view

[1] Supported by grants from the University Research Committee from funds provided by the Wisconsin Alumni Research Foundation and research grant M-1090 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. The authors wish to express their appreciation to Dr. Eric G. Heinemann for his suggestions in the preparation of the manuscript.

either the test object or the photometric field which is the same size and located in a comparable position to the test field. The test object is illuminated by a series of variable sources which are not visible to *S*. The luminance of the photometric field can be adjusted by *S* to match the test object under the various levels of illuminance employed.

For the continuous exposure duration, the apparatus and procedure were identical to that used in the previous study with the exception that the test object had a higher reflectance value (43% instead of 19.8%). Thirty-six volunteers from elementary psychology classes, who were unaware of the purpose of the study, were divided into three groups of equal size. All *S*s matched the photometric field with the test object at the following levels of illuminance: .0051, .047, .403, 25.4, and 1,522 ft.-candles. The procedure for the three groups differed only with respect to the reflectance of the background against which the stimulus was viewed. One group viewed the test object against a piece of "black" "flock" paper (reflectance 0.85%) 18 in. wide by 24 in. high on which the test object was centered. For the other two groups the backgrounds were a "gray" of the same reflectance as the stimulus or a "white" (reflectance 62%). The remainder of the chamber was painted "flat" black (1.78% reflectance). Half of the *S*s in each group were presented the various levels of luminance in increasing order, the other half in decreasing order. Six judgments utilizing the method of adjustment were made by each *S* at every luminance level. Between matches

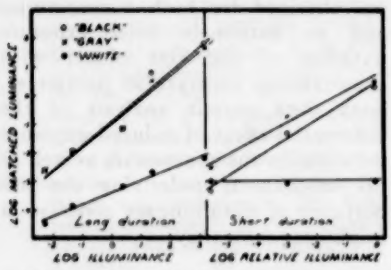

FIG. 1. Log matched luminance of the comparison field as a function of the log of the ambient illuminance in the test chamber for a continuous and a brief exposure duration and for three background conditions. On this plot a slope of zero represents a prediction in terms of the "law" of brightness constancy. A slope of unity represents the "additivity law of luminance." The "white" background has a reflectance of 62.00%, the "gray" 43.00%, and the "black" 0.85%.

the photometric field was not visible to *S*, and when it reappeared, it had been offset by *E*. No time limit was placed on *S*s, who required about an hour, including time for instructions and dark adaptation, to complete the experiment.

For the short duration the same stimulus conditions obtained as for the long duration except that the test object was viewed by the .0002-sec. duration flash produced by a gas discharge tube (Kemlite electronic flash tube, type 5W). The firing of the tube was controlled by *S* who could illuminate the stimulus chamber with the brief flash as often as every 5 sec. Between flashes, the photometric field was continuously visible while the location of the center of the test object in the dark stimulus chamber was indicated by a dimly illuminated fixation point. Thirty-six additional *S*s were divided into groups as for the long duration and each made eight judgments at each of the three illuminance levels. Illuminance level was controlled by placing neutral tint filters (made from photographic film) in front of the flash tube. Since the luminance produced by the short duration flash cannot be calibrated in the usual manner, an alternative procedure was employed. The maximum log luminance level available is arbitrarily specified as 0, and the other two levels expressed relative to the maximum as determined from the calibration values of the filters. The order of presentation of the luminance levels was determined by a 3×6 latin-square design with two *S*s assigned to each row. The *S* was allowed as many flashes as were required to make a satisfactory match after which the photometric field was extinguished and offset by *E*. A session consisting of dark adaptation, instructions, and 24 judgments, required about 1 hr.

RESULTS

The results expressed as the mean log luminance of the photometric field in foot-lamberts which matches the test field for the two durations, the various luminance levels, and the three background conditions, are presented in Table 1.[2] The data for the

[2] Individual data have been deposited with the American Documentation Institute. Order document No. 5270 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

MEAN LOG MATCHED LUMINANCE IN FOOT-LAMBERTS AS A FUNCTION OF ILLUMINANCE LEVEL WITH EXPOSURE DURATION AND BACKGROUND CONDITION AS PARAMETERS

Background	Continuous Exposure					Brief Exposure (.0002 sec.)		
	Log Ambient Illuminance, Foot-Candles					Log Relative Illuminance		
	− 2.292	− 1.328	0.395	1.405	3.182	− 5.60	− 3.00	0
"Black"	− 2.61	− 1.94	− 1.17	0.78	1.78	− 3.32	− 1.33	0.28
"Gray"	− 2.61	− 1.96	− 1.18	0.51	1.62	− 2.54	− 0.74	0.45
"White"	− 4.46	− 3.85	− 3.18	− 2.67	− 2.19	− 2.99	− 3.08	− 2.99

long exposure duration are plotted as a function of log test-object illuminance and for the short duration as a function of log relative illuminance in Fig. 1. On this plot a horizontal line represents a prediction in terms of the "law" of brightness constancy, a theoretical condition in which matched luminance remains constant despite variation in illuminance. The other theoretical extreme, the "additivity" law of luminance, for which any change in test-object illuminance should result in a corresponding change in the matched luminance, is represented by a line of unit slope and corresponds to a prediction based solely on energy relationships (**10**).

It will be observed in Fig. 1, that the background against which the stimulus is viewed is a determiner of the slope of the matched luminance-log illuminance function and, as well, of the extent to which the data are in accord with the "law" of brightness constancy. For the "black" and "gray" background conditions, the slopes of the best fitting straight lines as determined by the method of averages are .86 and .82, respectively, for the long exposure duration. With the introduction of test-object contrast by viewing against a "white" background, the slope is reduced in the direction of greater constancy to .42. In addition, it should be noted that the entire function is shifted to lower matched luminance values resulting from the inhibition, by simultaneous contrast, of the test field. These results agree with those of the previous investigation which produced slopes of .82, .67, and .42 for the "black," "gray," and "white" backgrounds (**7**).

For the short exposure duration, the various backgrounds produce the same relative effect on the matching functions obtained. For the "no contrast" conditions produced by the "black" and "gray" backgrounds, the tendency toward constancy is similar with slopes of .64 and .54, respectively. However, as for the long duration, the introduction of the contrast producing "white" background increases constancy and, in this case, produces matches which are in agreement with the "law" of brightness constancy, the computed slope being 0.00.

Thus, for both exposure conditions, there is some tendency toward constancy when the test object is viewed against a background of equal or less reflectance. However, when the magnitude of the test-object contrast is increased by viewing against a white background, the tendency toward brightness constancy also increases. Moreover, comparison of the functions for the long and short durations

reveals that, for comparable background conditions, there is a greater tendency toward brightness constancy for the short as compared with the long duration.

Discussion

It has been deomonstrated for a continuous exposure duration, in confirmation of our previous study, and also for an extremely brief exposure duration, that the introduction of simultaneous contrast by viewing a test object against a "lighter" background increases the tendency toward brightness constancy. These data support the previous conclusion that simultaneous contrast is an important condition mediating brightness constancy.

The additional finding of the present study is that reduction of exposure duration increases the tendency toward constancy as compared with the continuous exposure duration under the same viewing conditions. This may be interpreted in terms of the role of simultaneous contrast in brightness constancy if the contrast effect is differentially affected by luminance at the long and the short exposure durations. Comparison of the log matched luminance values at the highest illuminances used reveals that the matched luminance values are three to four times lower for the "white" background as compared with the "black" or the "gray" backgrounds. As illuminance level is lowered, however, this difference decreases so that at the lowest levels the matched values are practically identical for all background conditions at the short duration, and less than one and one-half log units apart for the long duration. Thus, the difference in matched luminances, and by inference also the magnitude of the contrast effect, decrease at a more rapid rate for the short as compared to the long duration as illuminance level is lowered. The result of this differential effect is that the slope of the "white" background, short duration, function is lowered more in the direction of brightness constancy than is the corresponding function for the continuous exposure.

The value of this analysis can be confirmed only by the presently unavailable functional dependencies of the simultaneous contrast effect on exposure duration at various luminance levels. However, experimental analysis of visual response in terms of time and intensity of stimulation is seen to be a heuristic approach to the problems of perceptual constancy. The observation of the present study, that a matching response remains invariant over an illuminance range of more than 100,000 to 1 under conditions of short exposure duration and simultaneous contrast, indicates the presence of an important mechanism subserving the constancy process.

Summary

Luminance matches were obtained between a "gray" test object viewed over a range of ambient illuminance values, three background conditions, and for two exposure durations.

All matches lie between theoretical predictions based on the "additivity" law of luminance and the "law" of brightness constancy. With increase in simultaneous contrast by viewing the test object against a background of higher reflectance, the matches exhibit a greater tendency toward brightness constancy. With decrease in exposure duration to .0002 sec., all functions lie closer to the prediction in terms of brightness constancy than for a continuous exposure duration.

The results are discussed in terms of the role of simultaneous contrast in brightness constancy.

References

1. Diamond, A. L. Foveal simultaneous brightness contrast as a function of inducing and test-field luminances. *J. exp. Psychol.*, 1953, **45**, 304–314.
2. Diamond, A. L. Foveal simultaneous contrast as a function of inducing-field area. *J. exp. Psychol.*, 1955, **50**, 144–152.
3. Graham, C. H., & Kemp, E. H. Brightness discrimination as a function of the duration of the increment in intensity. *J. gen. Physiol.*, 1938, **21**, 635–650.
4. Hartline, H. K. Intensity and duration in the excitation of single photoreceptor units. *J. cell. comp. Physiol.*, 1934, **5**, 229–247.

5. Heinemann, E. G. Simultaneous brightness induction as a function of inducing- and test-field luminances. *J. exp. Psychol.*, 1955, **50,** 89–96.
6. Leibowitz, H., Mote, F. A., & Thurlow, W. R. Simultaneous contrast as a function of separation between test and inducing fields. *J. exp. Psychol.*, 1953, **46,** 453–456.
7. Leibowitz, H., Myers, N. A., & Chinetti, P. The role of simultaneous contrast in brightness constancy. *J. exp. Psychol.*, 1955, **50,** 15–18.
8. Leibowitz, H., & Bourne, L. E. Time and intensity as determiners of perceived shape. *J. exp. Psychol.*, 1956, **51,** 277–281.
9. Leibowitz, H., Chinetti, P., & Sidowski, J. Exposure duration as a variable in perceptual constancy. *Science*, 1956, **123,** *668–669*.
10. Leibowitz, H. Relation between the Brunswik and Thouless ratios and functional relations in experimental investigations of perceived shape, size, and brightness. *Percept. mot. Skills*, 1956, **6,** 65–68.

(Received June 29, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

THE RELATIONSHIP OF PERFORMANCE LEVEL TO LEVEL OF AROUSAL [1]

RICHARD G. STENNETT [2]

Allan Memorial Institute of Psychiatry and McGill University

Over the past thirty years several authors (**2, 4, 7, 9**) have suggested that much of the confusion surrounding the concepts of emotion and motivation is conceptual rather than factual. Duffy (**5**), for example, has stated that behavior can be adequately described by categories more basic than the overlapping ones of motive and emotion. She suggests the dimension of "intensity" or "energy mobilization" and attempts to show that there is no discontinuity between motivated and emotional behavior evaluated in terms of this dimension. Freeman (**6**) and Hebb (**8**), although deriving their hypotheses from different theoretical contexts and using different descriptive terms, have made similar assertions. It would seem, therefore, that there is a growing emphasis upon a general dimension of behavior which is related to the intensity of organismic functioning and which, on a physiological level, puts meaning into the often-employed concept of "state of the *O*."

Both of these authors, on the basis of this conceptual revision, have suggested an important hypothesis with regard to the relationship between performance level and this intensitive dimension, which we prefer to call the *arousal continuum*, viz., that there is an inverted-**U** relationship between level of arousal and performance level.

The only evidence bearing directly upon this hypothesis is Freeman's report on a single case in which he employed reaction-time scores to measure level of performance and palmar conductance to measure arousal. The present study was undertaken to provide additional data in terms of which the validity of this hypothesis could be evaluated.

In light of the fact that several recent studies from this laboratory (**1, 10, 12**) have suggested that the level of tonic muscular tension developed during the performance of a motor task could be used as a measure of the degree of arousal, electromyographic (EMG) recordings were taken from four different muscle groups in addition to palmar conductance. An auditory tracking task was employed to provide scores which could be used as a measure of performance level.

METHOD

Design

To test this hypothesis a self-control, counterbalanced design was used in an experiment

[1] This paper represents part of a thesis (**11**) presented to the Faculty of Graduate Studies of McGill University for the degree of Doctor of Philosophy. This research was supported in part by the Research and Development Division, Office of the Surgeon General, Department of the U. S. Army, under Contract No. DA-49-007-MD-626, in part through a fellowship awarded by the National Research Council of Canada, and in part by the Defence Research Board of Canada under Grant No. 9425-04. The author is indebted to Dr. R. B. Malmo for his generous provision of laboratory facilities and supervision of the research reported here, and to Dr. J. F. Davis who designed and supervised the author's construction of the auditory tracking apparatus. The credit for the basic idea of auditory null-point tracking belongs to Mr. R. E. Quilter.

[2] Now at Ontario Hospital, St. Thomas, Ontario, Canada.

consisting of four different conditions:[2] (*a*) *Exertion* (*E*): This condition consisted of two, 2-min. "trials" at the beginning (E_1) and at the end (E_2) of the testing sequence in which *S*s were required to exert more rotational force on the controller knob of the tracking apparatus and hear more sound than they would during any actual tracking trial. This condition was inserted primarily to evaluate the effects of the simple physical requirements of tracking. (*b*) *Learning* (*L*): During this condition *S*s were given tracking trials until they reached arbitrary criteria of at least 10 trials and at least 3 trials without improvement. Encouragement from *E* and small financial rewards of 25 cents for bettering their best score were provided with the hope that the later stages of the learning condition would produce the optimal level of arousal for best performance. (*c*) *Calibration* (*C*): Four trials were presented to *S*s during this condition under the guise of their being necessary for the purpose of recalibrating the recording equipment. The *S*s were told that these trials were the same as those during the learning condition except that, because *E* was busy with calibration procedures, he could not keep track of their scores. It was hoped that the absence of knowledge about their performance, the withdrawal of the monetary incentive and encouragement by *E*, would produce a fall in arousal to a level below that of Cond. L. (*d*) *Incentive* (*I*): This condition, designed to produce above-optimal arousal, consisted of four trials in which the incentives offered to *S*s were changed. In the first of these trials, *S* was told that the best score anyone had made on the test so far was a certain figure (10% below the best error score *S* himself had made up to that point) and were encouraged to try to break that "record." In the following three trials *S*s 1–14 were given an opportunity to earn a bonus of $5.00 *each* time they equaled or bettered their best score; *S*s 15–31 could earn $2.00 extra and avoid a 150-v. shock by similar performance.

The order in which Cond. C and I were presented was alternated with successive *S*s so that there were, in effect, two different groups: one which had the conditions in the order E_1LCIE_2 and a second which had the conditions in the order E_1LICE_2. In this way the effects of the order of presentation of conditions could be evaluated.

The design combined the necessary conditions (L, I, C) to test the hypothesis with the controls required to: (*a*) partial out the effects of task demands (Cond. E), (*b*) eliminate performance score variation due to learning by practicing to a plateau, and (*c*) control for the influence of order of presentation of conditions (counterbalanced design).

[2] Other experimental conditions and the recording of other physiological variables (EEG, respiration rate, and heart rate) were also incorporated in this study. Since they bear little relationship to the hypothesis under test here, they will not be reported in this paper. The interested reader may consult the writer's doctoral work.

Subjects

The *S*s were 31 male, right-handed students at McGill University who came from various faculties and courses and whose ages ranged from 17 to 23. They were paid $1.00 per hour for their services.

Apparatus

Tracking apparatus.—The tracking apparatus was specially designed to present *S* with a continuously varying point-of-no-sound or null point which he could track by rotating a controller knob. On the left "side" of this null point *S* could hear a 800-cps tone which increased in amplitude as a direct, linear function of his deviation from the null point on that side. On the right "side" of the null point *S* could hear a 1000-cps tone which similarly increased in amplitude corresponding to his deviation from the null point to the right. Thus, by the use of these auditory cues, it was possible for *S* to orient himself with respect to the null point or "on course" position at all times.

Subject's task.—A "problem" generator provided movement of the position of the null point relative to the resting position of the controller knob according to a monophasic triangular function of time. Each tracking trial consisted of two 1-min. cycles of this function and required *S* (for correct performance) to turn the knob at a constant rate from its "on course" starting position to the left through an arc of 10°, back to the starting position, and then repeat this cycle. The controller knob, fitted with a strain-gauge for measuring grip pressure, was spring loaded so that 300 gm. of rotary force were required to turn it from the starting position to its point of maximum excursion.

Error measurement.—An amplifying and error-detecting device provided continuous monitoring of *S*'s tracking performance. A quantitative measure of his performance was immediately available at the end of each trial in the form of time-off-target scores which were read directly from a .01-sec. Standard Electric time clock.

Recording Technique

EMG.—Muscle potentials from the pronator quadratus and pronator teres muscles from the active (right) and passive (left) arms were recorded on the chart of a custom-built, ink-writing, Edin EMG. Electronic integrators (employing the condensor charge-discharge principle) attached in parallel across the galvanometers of this EMG unit, summated the muscle potentials over successive 4-sec. periods. A fifth Edin-integrator channel was used to record the output of the strain-gauge circuit of the controller knob and thus *S*'s grip pressure.

Conductance.—A simple Wheatstone bridge with sufficient voltage and series resistance to provide constant current through *S* (± 2% from 0–80 K) was used in conjunction with a standard decade resistance box for balancing the bridge. The circuit was designed so that, when balanced, the decade resistance reading was identical with *S*'s resistance. An Edin DC amplifier and recording galvanometer were used as a null-indicating device and changes in *S*'s resistance were measured as unbalance of this null reading (1000 ohms change = 4.5 mv. unbalance) on the tape of an auxiliary chart drive. Calibration was arranged so that each 3 mm. of pen deflection represented 1000 ohms resistance change.

Electrodes.—Bipolar sponge electrodes, soaked in electrode jelly and attached to *S* by lastonet bands, were used for all EMG placements and for the administration of a 100–150-v. condenser shock to the calf of *S*'s right leg. Typical silver-silver chloride electrodes of approximately 1 sq. cm. in size were used for recording resistance changes. They were inserted into a 1-in. cube of cellulose sponge which was soaked in saline. These sponge-covered electrodes were placed in the palm and dorsum of *S*'s left hand, covered with 2-in.-square pieces of thin rubber (to prevent evaporation of the saline), and held in place by a lastonet band.

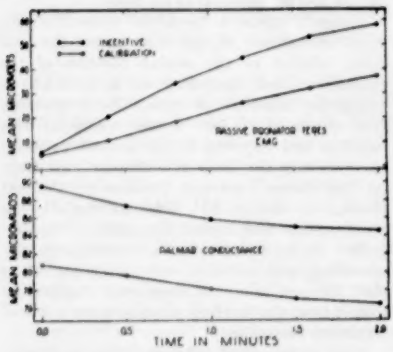

FIG. 1. The within-trial time course of passive pronator teres EMG and palmar conductance for the incentive and calibration conditions. Since there were four incentive and four calibration trials for each of 31 *S*s, each point plotted is the mean of 124 observations.

Method.—After the tracking task had been explained to *S* (by analogy with a blind-flying situation), he was seated in a semireclining chair and all of the electrodes were attached and checked according to the standard procedure employed in this laboratory (**3**).

The pronator muscles of *S*'s left arm were maintained in a position of slight stretch by the use of a positioning device, and an approximately isometric condition of the "active" pronators was ensured by the combination of positioning devices and orientation of the controller knob. A short sample of record was taken before the beginning of each 2-hr. recording session to ensure that no artifacts were present. The experimental conditions outlined above were then administered.

Measurement of Records

All data from the four EMG placements, the grip-pressure channel, and palmar conductance were measured for all trials for all *S*s. However, since optimal arousal was defined operationally for each *S* as the average of the conductance values corresponding to his four best performance scores in the learning condition, the remaining data from the learning condition will not be considered in detail.

"Active" EMG and grip pressure.—For each trial for each *S*, six integrator waves were measured. These waves represented activity during the following 4-sec. intervals: 12–8 and 4–0 sec. before the beginning of the trial (preperiod); 12–16, 44–48, 72–76, and 104–108 sec. during the trial; and 4–8 sec. after the trial (postperiod). By sampling in this manner, within-trial measurements were made at points in which the physical requirements of the task were equal.

"Passive" EMG.—The absence of task-related activity permitted more extensive sampling of these two channels. Integrator waves representing activity in the following 4-sec. intervals were measured: 12–8 and 4-0 sec. of the preperiod; 20–24, 44–48, 68–72, 92–96, and 116–120 sec. during each trial; and 4–8 sec. of the postperiod.

Palmar conductance.—Measurements of resistance level were made (to the nearest 333 ohms, ensuring that apparatus error was less than the unit of measurement) at the beginning

TABLE 1

WITHIN-TRIAL CHANGES IN PALMAR CONDUCTANCE AND THE PASSIVE PRONATOR TERES EMG

Trial		Mean Conductance in Micromhos			Mean Level of EMG in Microvolts		
		First 1 Min.	Second 1 Min.	Diff.	First 4 Sec.	Last 4 Sec.	Diff.
Learning	1	87.3	84.7	− 2.6	5.50	40.26	+ 34.76
	2	85.9	83.3	− 2.6	2.19	41.45	+ 39.26
	3	85.4	82.9	− 2.5	3.26	34.73	+ 31.47
	4	85.8	83.8	− 2.0	3.30	43.80	+ 40.50
	5	84.5	82.5	− 2.0	5.50	56.12	+ 50.62
	6	83.5	82.8	− 0.7	2.19	51.83	+ 49.64
	7	84.1	83.1	− 1.0	2.80	49.83	+ 46.03
	8	83.3	81.3	− 2.0	4.77	52.19	+ 47.42
	9	84.0	81.8	− 2.2	3.80	48.19	+ 44.39
	10	85.8	83.7	− 2.1	3.35	48.96	+ 45.61
Calibration	1	82.8	79.5	− 3.3	3.97	42.03	+ 38.06
	2	80.5	77.5	− 3.0	3.50	32.30	+ 28.80
	3	77.6	73.9	− 3.7	5.31	33.21	+ 27.90
	4	79.9	77.0	− 2.9	4.64	41.91	+ 37.27
Incentive	1	81.9	79.0	− 2.9	4.47	58.40	+ 53.93
	2	88.7	86.0	− 2.7	4.90	64.26	+ 59.34
	3	89.7	86.0	− 3.7	5.07	55.63	+ 50.56
	4	90.8	87.5	− 3.3	8.20	55.66	+ 47.46

and end of each trial and every 15 sec. during the trial, i.e., nine measurements were made for each *S* for each trial. The resistance values were converted into micromhos and all calculations were done on these transformed values.

RESULTS

Within-trial changes.—The within-trial time courses of palmar conductance and EMG for the passive pronator teres muscle, for Cond. I and C, are shown in Fig. 1. Each curve is an average of the four trials for all *S*s for each of these conditions. Table 1 presents the within-trial change in these two variables for all of the tracking trials. These results indicate that there was no exception to the general rule that from the beginning to end of a trial EMG rises whereas palmar conductance falls. This is in apparent contradiction to the trend that, in terms of trial-to-trial changes, the two variables seem to be positively correlated.

Interaction effect.—Table 2 presents the conductance means calculated separately for *S*s who had the experimental conditions in the order Learning, Incentive, Calibration (Group LIC), and for those who had the conditions in the order Learning, Calibration, Incentive (Group LCI). A pronounced interaction effect between conductance level and order of presentation of the conditions is present. This means that during Cond. C, *S*s in Group LIC did not show a fall in conductance which was significantly below that of the L or optimal condition. (For purposes of convenience, we will use the symbol L to refer to "optimal" or four best performance trials in the learning condition.) Similarly, during Cond. I, *S*s in Group LCI did not show a rise in conductance to a level significantly above that of the optimal condition. A statistical test of this

TABLE 2

COMPARISON OF MEAN CONDUCTANCE VALUES FOR LCI AND LIC GROUPS

Comparison	Group LCI (N = 16)			Group LIC (N = 15)		
	Means	Diff.	P	Means	Diff.	P
Optimal-Calibration	83.82–77.76	6.06	< .001	78.33–79.54	− 1.21	N.S.*
Incentive-Optimal	89.18–83.82	5.36	> .10	82.78–78.33	4.45	< .05
Incentive-Calibration	89.18–77.64	11.54	< .001	82.78–79.54	3.24	< .001

* Since, in this comparison, the mean conductance during the calibration condition was higher than that for the optimal condition, no t test was made.

interaction effect was made by comparing the conductance means for Cond. L and C for Group LIC and the means of Cond. L and I for Group LCI. In both cases the interaction effect, as expected, produced no significant differences. However, all other differences expected from the effects of the experimental conditions were in the direction predicted and highly significant (Table 2).

This interaction effect dictates that, in testing our hypothesis regarding the relationship of arousal level to performance level, we should compare the conductance values and performance scores for the "low arousal" (C) and "optimal arousal" (L) condition for Group LCI only. Conversely, for the comparison of "optimal arousal" and "high arousal" (I) conditions, we should compare the conductance values and performance scores of only Group LIC.

Conductance level and error scores.—For the purpose of experimental test and in light of the interaction effect the hypothesis that an inverted-U relationship exists between the level of arousal and performance level can be restated operationally in terms of four propositions. If we assume that the average of the conductance values corresponding to the best four scores in the learning condition represents the best estimate of the "optimal" level of arousal then: (*a*) In Group LCI the average of the performance scores for Cond. C should be worse than the average of the best four scores in Cond. L. (*b*) In Group LIC the average of the performance scores for Cond. I should be worse than the average of the four best scores in Cond. L. (*c*) In Group LCI the average of the conductance values for Cond. C should be lower than the average of the conductance values corresponding to the four best trials in Cond. L. (*d*) In Group LIC the average of the conductance values for Cond. I should be higher than the average of the conductance values corresponding to the four best scores in Cond. L.

Table 3 shows the means and P values for each of these comparisons. In all cases one-tailed t tests were applied. All of these results, then, were significant and in the direction predicted by the hypothesis.

EMG and error scores.—Since testing the hypothesis that the level of arousal (measured in terms of EMG tension) bears an inverted-U relationship to level of performance (measured in terms of errors) involves the same considerations as those in testing the relationship of conductance level to error scores, the same procedure was employed except that EMG values representing the total amount of EMG tension recorded from the passive pronator teres muscle

TABLE 3

COMPARISON OF CONDUCTANCE AND PASSIVE PRONATOR TERES EMG MEANS WITH ERROR SCORE MEANS FOR THE LCI AND LIC GROUPS

Measure	Group LCI (N = 16)				Group LIC (N = 15)			
	Calibration	Optimal*	Diff.	P	Incentive	Optimal*	Diff.	P
Error Score	10.31	6.55	3.66	< .001	7.69	5.56	2.13	< .02
Conductance	77.80	83.80	6.00	< .01	82.80	78.30	4.50	< .05
Passive Pronator Teres EMG's**	129.46	168.60	39.14	< .01	210.06	121.60	88.46	< .001

* Optimal trials are those four trials in the learning condition in which S made his best scores.
** The values, upon which these means are based, were derived by simply summating the five within-trial measurements. They are, therefore, "total tension" estimates rather than average trial values.

within each trial were substituted for average conductance values.

Table 3 shows the means and P values for each of the comparisons involved. In all cases one-tailed t tests were applied. These results, then, are also significant and in the direction predicted by our hypothesis.

Further EMG results.—Figure 2 shows the within-trial changes in the active pronator teres and active pronator quadratus muscles, and grip pressure for all Ss for the four incentive, four optimal, four calibration, and four exertion trials. Points plotted are those where the physical requirements of the task were equal. Within-trial changes in the passive pronator teres and quadratus muscles are shown in Fig. 3. The absence of task-demand activity from these EMG placements permits plotting six points instead of four.

From these two figures it should be noted that, in terms of level, the curves for all five variables are ordered identically with respect to experimental conditions. This means that the results of the test of the relationship between EMG tension and error scores would be invariant with respect to the particular muscle selected for study. A second point to note is that the curves of grip pressure fall from the beginning to the end of the trial whereas all of the EMG curves rise. A final point of interest is that the curves for all variables are lowest during the exer-

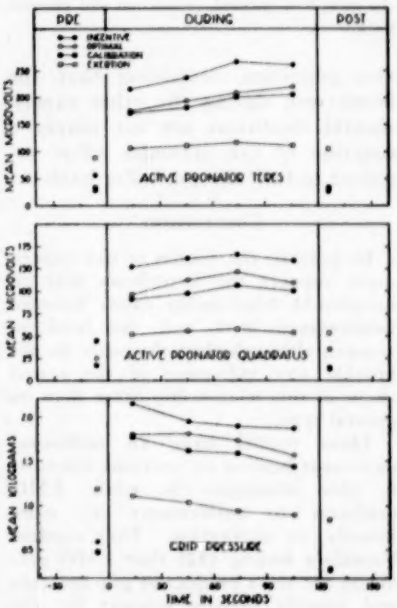

FIG. 2. Within-trial changes in the active pronator teres and quadratus muscles and grip pressure for the different experimental conditions.

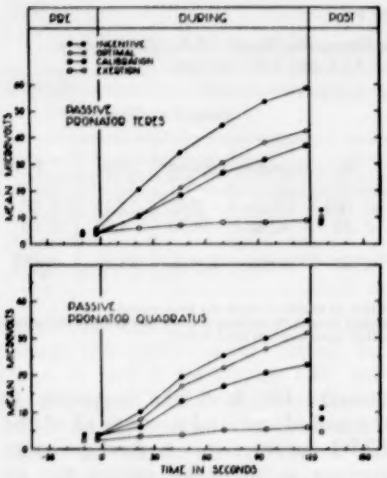

FIG. 3. Within-trial changes in the passive pronator teres and quadratus muscles during the incentive, optimal, calibration and exertion conditions.

tion condition, indicating that the levels seen during the other experimental conditions are not simply a function of the muscular effort required to turn the controller knob.

DISCUSSION

In general, the results of this experiment support the hypothesis that an inverted-U relationship exists between performance level and the level of arousal, although they obviously do not provide any indication of the actual shape of this relationship other than its general type.

These results have an additional important bearing on previous attempts in this laboratory to relate EMG gradients to performance or, more broadly, to motivation. They confirm Surwillo's finding (12) that EMG gradients are not a product of grip pressure and provide definite support for the hypothesis that incentive is the most important experimental variable concerned with the steepness of these gradients. However, a more complete picture is provided if the EMG gradient phenomenon is considered in relation to the broader dimension of arousal. In terms of this concept it is possible to understand why, for example, the relationship discovered by Bartoshuk (1) in mirror tracing, and confirmed by Surwillo (12) in visual tracking, between performance scores and EMG gradient steepness did not reach higher levels of statistical significance. It seems very likely that some of their *S*s were at a point on the arousal continuum where further increases in arousal due to the introduction of incentive would lead them into the range where performance level and arousal are negatively correlated. Thus, in the small samples used (*N*'s of 10–20), two or three *S*s operating at above-optimal arousal levels could reduce the significance of the correlations considerably.

SUMMARY

A test of the hypothesis that an inverted-U relationship exists between the level of arousal and performance level was made by comparing the performance of 31 *S*s on an auditory tracking task under different conditions of incentive. These conditions ranged widely from one in which *S* was under the impression that his scores were not even being recorded to one in which his score determined whether or not he avoided a 100–150-v. shock and earned bonus money of from $2.00 to $5.00. When the effects of learning on performance scores were controlled and an interaction effect of the order of presentation of the experimental conditions had been partialled out, the data of this study gave strong support to the hypothesis. The hypothesis held regardless of whether palmar conductance level or the EMG response of any one of four different muscle groups was used as the criterion of arousal.

REFERENCES

1. BARTOSHUK, A. K. Electromyographic gradients as indicants of motivation. *Canad. J. Psychol.*, 1955, **9**, 215–230.
2. BINDRA, D. Organization in emotional and motivated behavior. *Canad. J. Psychol.*, 1955, **9**, 161–167.
3. DAVIS, J. F. *Manual of surface electromyography*. Montreal: Laboratory for Psychological Studies, Allan Memorial Institute of Psychiatry, 1952 (mimeo.).

4. Duffy, E. Emotion: an example of the need for reorientation in psychology. *Psychol. Rev.*, 1934, **41**, 184–198.
5. Duffy, E. The concept of energy mobilization. *Psychol. Rev.*, 1951, **58**, 30–40.
6. Freeman, G. L. The relationship between performance level and bodily activity level. *J. exp. Psychol.*, 1940, **26**, 602–608.
7. Freeman, G. L. *The energetics of human behavior.* Ithaca, N. Y.: Cornell Univer. Press, 1948.
8. Hebb, D. O. Drives and the C.N.S. (Conceptual Nervous System). *Psychol. Rev.*, 1955, **62**, 243–253.
9. Meyer, M. F. That whale among the fishes—the theory of emotions. *Psychol. Rev.*, 1933, **40**, 292–300.
10. Smith, A. A. An electromyographic study of tension in interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 32–36.
11. Stennett, R. G. The arousal continuum. Unpublished doctor's dissertation, McGill Univer., 1956.
12. Surwillo, W. W. Psychological factors in muscle action potentials: EMG gradients. *J. exp. Psychol.*, 1956, **52**, 263–272.

(Received June 29, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

LEARNING THROUGH STIMULUS SATIATION

M. RAY DENNY[1]

Michigan State University

The present study is related to the experiments on spontaneous alternation in the rat (**1, 3, 4, 5, 6, 8, 9**). As in most of these studies, a T maze with dissimilar arms is used. But, unlike the typical spontaneous alternation study, the rat is not permitted to respond freely. Instead, the rat is frequently forced so that in each block of six trials it goes twice as often to one side of the T as the other. The trials are well spaced and visits to either arm are always rewarded. Thus the rat is also rewarded twice as often on one side as the other. On free trials one might expect, therefore, that the rat would learn, at least early in training, to go where it is rewarded more often. The answer seems to be, however, that the opposite is true. On free trials the rat learns to go to the side it visits *less* often. One possible interpretation of this finding, as proposed here, is that conditioned inhibition or avoidance learning can be mediated by stimulus satiation effects and thus *S* learns to avoid the more satiated side. Another way of saying this is that *S* learns to respond so as to maximize stimulus variation.

METHOD

Sixteen hooded and albino female rats from 90–110 days of age from the colony of the Psychology Department of Michigan State University were employed. The apparatus was an enclosed T maze constructed of unpainted wood. The over-all dimensions of the arms and stem were 44 in. and 24 in., respectively. The alleys were 4 in. wide and 6 in. high and covered with hardware cloth. The end boxes and starting box were 10 in. long. The left arm and end box were lined with black posterboard while the right side was natural wood like the remainder of the maze. Doors were located at the entrances to the stem, each arm, and each goal box. The two doors on the left were painted black. To avoid tail pinching the doors at the choice point closed by coming up through a slot in the floor.

All *Ss* were habituated for a week, and placed on a food regimen of 9 gm. a day which included Calf Manna "reward" pellets. Each *S* was then given two trials a day in the maze with a 30-min. intertrial interval. The trials were administered in 3-day blocks of six trials. For half the *Ss*, four trials of each block of six were to the right or natural side and two were to the left or black side; for the other half the ratio was reversed. On Day 1 of the block the initial trial was always free and the second was always forced. Half of the time *Ss* which chose the more frequently visited side on the free trial were forced the same way, and half the time the opposite way. The *Ss* which chose the less often visited side were always forced in the opposite direction. On Day 2 the first trial was always forced and always in such a way as to make the ratio 2:1 in the desired direction. The second trial was always free. On Day 3 the first trial was either free or forced, depending upon *S*'s previous responses; and the second trial was always forced.

This block of six trials was repeated eight times. Then on the following day *Ss* were given two free trials separated by a 30-min. interval. A week later two more free trials were given with the same 30-min. interval. On every trial, regardless of the side visited, *S* received a .15 gm. pellet of Calf Manna.

RESULTS

The results are presented in Fig. 1. On both the initial free trial of Day 1 and the second free trial of Day 2 of each block, there was a progressive increase in the number of responses to the side less often visited, to a level of 88% and 94%, respectively. When two free trials were given in the same day the 88% level was still main-

[1] The writer is indebted to Elizabeth Matus and Elfriede Lange for running some of the *Ss*.

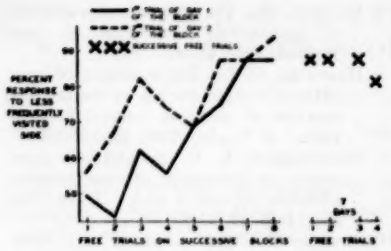

FIG. 1. The curves represent the growth of preference for the side of the maze which is less often visited when either the initial trial of the day is free or the second trial of the day is free.

tained, and even a week later *S*s were performing with a marked preference for the side which was previously visited less frequently.

The time for *S* to leave the starting box, run the maze and enter the goal box was recorded on each trial and these data suggest that the preference was mediated by an avoidance of the more frequently visited side. The running times to this side after the first few days were consistently greater than for the less often visited side. The time decreased with successive trials but then began to increase so that on the last two blocks, when the preference was well established, the mean running time was 38.8 sec. as compared with 16.2 sec. to the less frequently visited side. The difference yields a t of 2.4 ($P = .05$).

DISCUSSION

That the preference to the less frequently visited side was learned is supported by three independent lines of evidence: (*a*) The preference did not appear full blown at the start; a curve resembling a typical learning curve was obtained for the development of this preference. (*b*) At the end of the eight blocks of trials, when *S* had the opportunity on the second free trial to avoid the alley *last* visited, it persisted in preferring the alley less frequently visited. The point is that if stimulus satiation effects alone were producing the results, we would expect *S* to alternate. (*c*) There was retention of the preference on both free trials for at least a week. It should also be noted that results similar to those obtained here have been reported by Rothkopf and Zeaman (7).

In the present experiment, the concept of reactive inhibition appears less appropriate as an explanatory concept than stimulus satiation because the time between trials (30 min. and 24 hr.) is extremely long. In fact, when the arms are dissimilar the preference appears, though not so prominently, even when only one trial a day is given (9).[3]

The main conclusion is that avoidance learning can be mediated by stimulus satiation effects. Although such a finding is difficult to interpret in the framework of traditional Hullian reinforcement theory, it is consistent with the present writer's theoretical formulations, which include a stimulus satiation postulate (2). According to this position the responses prepotently elicited in a situation are the ones that become conditioned to this situation. In the present instance an avoidance reaction is assumed to be consistently elicited by the more frequently experienced stimuli and thus conditioned to them.

SUMMARY

Sixteen hungry, habituated rats were run in a T maze with differential arms (black and natural). For half of the *S*s, four trials of each block of six were to the left and two were to the right; for the others, the reverse was true. This was accomplished by a planned arrangement of free and forced trials within blocks of six trials. The *S*s were given two trials per day with an intertrial interval of 30 min. and food reward on every trial. At the end of eight blocks, *S*s were given two free trials on the next day, followed a week later by two more free trials.

[3] Exploratory research has indicated that when the arms *are similar* the tendency to go to the less often visited side is quite unstable, even for short intertrial intervals (15 min.).

On both the free Trial 1 of Day 1 and the free Trial 2 of Day 2 of each block, there was a progressive increase in the number of responses to the side *less often* visited, to a level of 88% and 94%, respectively. Correlated with this was an increase in running time on the forced trials to the other side. The preference for the less frequently visited side was retained for at least one week. It is concluded that the preference for the less frequently visited side was learned.

References

1. Dennis, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effects. *J. comp. Psychol.*, 1939, **28,** 305–312.
2. Denny, M. R., & Adelman, H. M. Elicitation theory: I. An analysis of two typical learning situations. *Psychol. Rev.*, 1955, **62,** 290–296.
3. Estes, W. K., & Schoeffler, W. S. Analysis of variables influencing alternation after forced trials. *J. comp. physiol. Psychol.*, 1955, **48,** 357–362.
4. Glanzer, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45,** 387–393.
5. Heathers, G. L. The avoidance of repetition of a maze reaction in the rat as a function of the time interval between trials. *J. Psychol.*, 1940, **10,** 359–380.
6. Montgomery, K. C. Spontaneous alternation as a function of time between trials and amount of work. *J. exp. Psychol.*, 1951, **42,** 82–93.
7. Rothkopf, E. Z., & Zeaman, D. Some stimulus controls of alternation behavior. *J. Psychol.*, 1952, **34,** 235–255.
8. Walker, E. L., Dember, W. N., Earl, R. W., & Karoly, A. J. Choice alternation: I. Stimulus vs. place vs. response. *J. comp. physiol. Psychol.*, 1955, **48,** 19–24.
9. Wells, R. H. Learning through stimulus satiation: one trial a day. Unpublished doctor's dissertation, Michigan State Univer., 1956.
10. Zeaman, D., & House, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, **41,** 177–186.

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

FACTORS IN THE RECOVERY FROM APPROACH–AVOIDANCE CONFLICT[1]

MITCHELL M. BERKUN[2]

Yale University

Conflict between incompatible responses in the rat is studied not only for its relation to conflicting motivation in man, but also for the knowledge it provides about fear and fear reduction. An approach habit is set up by reinforcing a hungry rat with food for running to one distinctive end of an otherwise homogeneous alley. A competing avoidance habit is then set up by giving the rat a painful electric shock while he eats. In subsequent trials a compromise response is observed—the rat goes only part way to the goal. Presumably fear extinguishes with successive nonshock trials, reducing the strength of the avoidance response and thus permitting the reappearance of approach behavior.

Miller and Kraeling (**3**) found that the proportion reaching the goal on the seventh postshock trial increased if the last four trials are given in a different alley. This finding was extended by Murray and Berkun (**5**), who provided rats recovering from conflict with a choice of alleys. After a brief exploration of the original alley, their rats turned to the most different alley, where they made the goal response. Only after recovery in the displaced alleys did they attempt (and accomplish) the goal response in the original alley.

What remains unknown is the contribution of this displaced experience to the recovery of the original response. The present experiment seeks to determine this by comparing the effect of continued trials in the original alley with the effect of continued trials in a different (or displaced) alley upon subsequent recovery in the original situation.

Displaced trials were given until a goal response occurred. This led to a second question for empirical test: Should this displaced response be reinforced with food?

Theoretical considerations implied an advantage for reinforcement, partly based on prevention of extinction of the approach, and partly on counterconditioning to reduce the avoidance. To investigate these alternatives, a third variable was introduced: intermittent (partial) reinforcement during acquisition of approach. If preventing extinction of approach were an appreciable part of the effect of the reinforcement, then this effect would be expected to interact with previous intermittent reinforcement. There is no known reason for intermittent reinforcement

[1] This paper is based on a dissertation presented to the Graduate School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is pleased to acknowledge with thanks the help and encouragement of his committee, Prof. Frank A. Logan and Prof. Fred D. Sheffield, and particularly his chairman, Prof. Neal E. Miller. Appreciation is also due Dr. Edward J. Murray, with whom this topic has often been discussed. This research was supported by a grant, M647, to Prof. Miller from the National Institute of Mental Health, National Institutes of Health, Public Health Service. More complete data and supplementary analyses are presented in the dissertation itself, available on interlibrary loan from Sterling Library, Yale University.

[2] Now at U. S. Army Leadership Human Research Unit, (HumRRO), Fort Ord, California.

to interact with counterconditioning. In addition to its possible interaction with the presence or absence of reinforcement during retraining, the question of the possible effects of intermittent reinforcement on resistance to conflict is also of interest.

Method

Apparatus and scoring.—The apparatus consisted of three wooden alleys placed side by side on a table, each 60 in. long, 6 in. high, with a hardware cloth top. One alley was 3 in. wide, with sides, ends, and floor painted flat black. The second was gray throughout, 5 in. wide; the third was white and 7 in. wide.

The reinforcement for each trial was a single 40-mgm. lab chow pellet presented on a platform jutting 2 in. into the end of the alley, ½ in. up from the floor. Each platform was surrounded by a hinged metal panel which *S* pushed back to expose the food. The platforms were 1, 2, and 3 in. wide, and painted black, gray, and white, respectively, to match the alley. A network of fine lines was scratched through the paint on the panels and on the galvanized metal floors of the alleys. A rat could then be shocked when he touched the panel with his nose to complete a circuit between panel and floor. A timing circuit was closed automatically by the weight of the rat on a hinged portion of the floor at the start end of the alley. This started a Springfield Electric clock, which then stopped when a rat moved the food panel about 2 mm.

Over the table top was mounted a cheesecloth tent, dimly lighted from within. With the room darkened, this served as a one-way screen to reduce visual cues from *E*'s movements.

Having placed *S* at the start end of the alley, *E* immediately moved to the goal end of the alley, and standing at the end of the table, looked down to observe *S*. A scale of 4-in. segments was marked off on the table alongside each alley. The first two segments, occupied by *S* in starting position, were labeled 0; the rest were numbered consecutively 1 through 13. The position of *S*'s nose could be noted, and the farthest segment which it entered during a trial was recorded as his distance score for that trial. In almost all cases *S* retreated to the start of the alley after having made his farthest excursion.

Subjects and design.—The *S*s were 64 male albino rats of the Sprague-Dawley strain, approximately 100 days old at the start of the experiment. They were housed in eight group cages, each cage containing a complete replication constituting one row of a latin square. The columns of the latin square gave the order in which the eight treatments were applied to successively run rats in each replication. Rats were pre-assigned at random to the cells of the square.

Alternate replications were run with the black alley as the original alley in which approach training was given and conflict induced. In these blocks, the displaced rats were given their first test trials in the wide, white alley. For the remaining rats, the white alley was the original alley. For both cases, the gray alley served as an intermediate phase in the recovery from displaced to original alley.

Procedure.—After seven days of habituation to daily 30-min. access to Purina Lab Chow Checkers (with water continuously available), approach training was begun. On Day 1, each *S* received two trials, with the panel propped open to expose the food. On Day 2 he was given 10 consecutive reinforced trials, about 10 min. apart. For the first four trials the panel was again held open to expose the food, but thereafter *S* was required to push it back unaided. On each of the next five days, 10 trials were given, five of which were nonreinforced for half the rats (the IR group). These five trials were assigned at random, with these restrictions: (*a*) both Trials 1 and 10 were to be reinforced, and (*b*) no more than two successive trials got the same condition. The pattern of reinforcement varied among days and among replications.

Since running speeds failed to increase from the fourth day to the fifth, or during the fifth day, it was assumed that learning had approached an asymptote. On the next day following, each *S* was first given a reinforced warm-up trial. On the next two trials immediately thereafter, shock was applied through the food shield. The 600-v. output of a radio transformer was reduced by a potentiometer circuit to 90 v. for the first shock trial and 135 v. for the second, delivered through a .25-megohm series resistor. After all *S*s had received the two shocks given on two successive trials, the voltage was raised to 400 for the third and final shock, and the .25-megohm resistor was removed from the circuit. Food was in the tray all through the shock series, but no rat withstood the shock long enough to get the food.

Some *S*s did not receive shock on their third trial because the avoidance induced by the first two shocks prevented them from reaching the goal. They were removed from the alley after 1 min. and subsequently run again. If necessary, *S*s were given up to 15 1-min. trials on this day, to receive the third shock. Rats which required more than 10 such trials were not given

any recovery tests until the next day. With this exception, recovery tests began at once.

Half of the rats began recovery in the original alley, half in the most different alley. After a goal response occurred in the different alley, *S* was given consecutive trials in the intermediate alley, until he made a goal response there. Then he was run to recovery in the original alley.

The *S*s were given up to 10 successive 2-min. trials, with from 15 to 30 min. between trials, the elapsed interval being independent of the treatment. If a goal response occurred in less than 2 min., the trial was ended by removing *S* immediately. Trials were continued until each *S* made three goal responses in the original alley (a total of five goal responses for the displaced group). For half of the *S*s the food tray was empty until after the second postshock goal response was accomplished.

A goal response consisted of eating the food, if food was available, or of pushing the panel far enough back to clearly expose to *S*'s view the place where food would have been (much farther than required to break the timing circuit). It is *E*'s observation that whenever a rat that had food available pushed the shield back to this visual criterion, he invariably ate the food at once. This procedure is conceivably subject to error resulting from a variable time between stopping the clock and eating the food. This is not likely to be significant, however, because the shock came from the shield and not the food. Thus, if *S* touched the shield at all, there is little residual avoidance to inhibit eating.

Animals not finishing their course of recovery in 10 trials were fed as usual, one-half hour after finishing the day's work. They were given 10 trials per day on succeeding days until recovery was complete. But feeding was omitted on the second and subsequent test days, in order to increase the approach tendency and thus bring in the recalcitrant *S*s.

The role of olfactory cues during those trials when reinforcement was withheld was, to some extent, controlled by maintaining a supply of pellets in an open jar just outside the end of each alley and by the fact that crumbs of lab chow always remained in inaccessible corners of the platform on which the pellets were presented.

Results and Discussion

Statements as to the reliability of the results which follow are all based on both tails of the distribution of a nonparametric statistic. Where data are in the form of proportions χ^2 was used; otherwise the rank-sum test (sometimes known as the Mann-Whitney *U* test) was used.

Equality of Randomized Groups

All *S*s went to the food tray and received the second mild shock on the trial immediately following the first shock. Variance was introduced in the number of trials required to recover from the second shock and receive the third shock.

Since both the original and the displaced groups were still working in the original alley, they should not have performed differently. There was no statistical evidence for a difference between these groups in speed during training, in the number of trials required to recover from the second shock, or in the distance traveled on the first trial following the second shock. Similarly, the groups subsequently to be fed or not fed, respectively, on the first two goal responses following the third shock, were equal in training speed and in both measures of rate of recovery from the second shock.

After the third and strongest shock, the course of recovery of the goal response was observed by recording (*a*) the farthest distance traveled on each successive 2-min. trial until a goal response occurred, (*b*) the number of such trials, and (*c*) the time required for each goal response which did occur.

General Pattern of Recovery

Although only five *S*s responded by operating the goal device on the first trial after the third shock, all 64 *S*s recovered by Trial 58. Half of the *S*s were first run in the displaced alley, then were run to a goal response in the intermediate (gray) alley, then back in the original alley. The rest

of the rats worked in the original alley only.

For all 64 *Ss*, the mean number of trials required for recovery of the first goal response, regardless of the alley in which it occurred, was 11.0. For the 32 *Ss* recovering in the original alley, the mean number of trials required is 12.6. This is reliably greater than the mean of 2.5 trials required for recovery from the second shock ($P < .001$).

The size of this effect of the third shock can be reduced if a criterion less strict than that of a complete goal response is used for the third shock. More comparable to the criterion of recovery from the second shock (i.e., just brushing the food panel) is the number of trials after the third shock until the first entrance into the final 4-in. segment of the alley. The 32 original-only *Ss* require a mean of 6.2 more trials to achive this criterion than they required to recover from the second shock, a difference still reliable at well past the .001 level. Likewise, the distance traveled in the original alley immediately after the third shock is shorter than that traveled after the second shock ($P < .001$). This stronger avoidance after the third shock would be expected on two bases: (*a*) the increase in the strength of the shock, and (*b*) the fact that *Ss* have received an additional shock.

Displaced Performance vs. Performance in the Original Alley

Distance.—On each trial, a score from 0 through 13 was assigned for the farthest distance from the start that *S* traveled, whether or not a goal response occurred. The curves of recovery of distance after the third shock are presented in Fig. 1. It can be seen that the displaced *Ss* start out by going farther down the different alley than the original-only *Ss* go down the original alley. On the first trial this difference is reliable at well beyond the .001 level; for the first block of five trials it is reliable at the .002 level. This confirms the effect of displacement upon distance traveled as previously reported by Murray and Berkun (**5**).

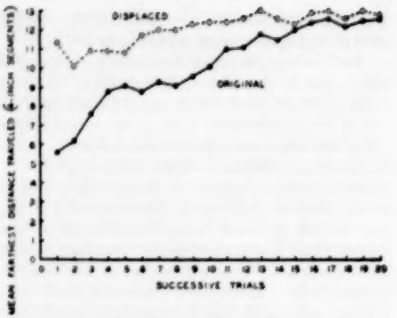

FIG. 1. Recovery of approach by displaced rats in the different alley and original rats in the original alley. Each mean is computed for 32 *Ss*, on the assumption that *Ss* who have made a goal response would have earned the maximum score of 13 on all subsequent trials.

It is also apparent that distance traveled increases with successive trials. The increment in distance from the first block of five trials to the fourth is reliable for the displaced group at the .02 level and for the original group at far beyond the .001 level. This recovery of approach is as predicted from the hypothesis that fear tends to extinguish with successive nonshock trials in the apparatus.

A good part of the apparently gradual increase in distance as a function of trials can be accounted for by the all-or-none shifting of an increasingly larger proportion of *Ss* from the start to the goal end of the alley. To illustrate this: 25 of the displaced *Ss* and 12 of the original *Ss* first passed Distance 11 of the alley on the same trial on which they first

passed Distance 5, indicating that in the middle of this short alley a majority of rats do *not* add regular increments in distance traveled per trial. Near the goal, however, there is decidedly less of this all-or-nothing type of progression. The *S*s do not go completely up to the food as soon as they finally start to run up, but instead spend an average of five trials (greater than zero at $P < .001$) sniffing the food panel before they finally operate it.

Reversals in the progress of recovery seem to be insignificant. Of the 52 *S*s who scored 13 for a trial before ever making the goal response, only 18 subsequently regressed to a distance score of 12 or less. After the first original goal response, 5 *S*s of 64 regressed to just sniffing in Segment 13 before completing the series of three original goal responses.

Trials.—It will be recalled that, after the third shock, the displaced *S*s were run in the displaced alley until they recovered the goal response there, and were then run in the intermediate (gray) alley until they responded there also. Then they were permitted to recover the original goal response. The mean numbers of 2-min. trials required by these *S*s up to and including the first goal response in each of the three alleys were, respectively, 9.4, 3.0, and 2.6. These total 15.0 trials required by the displaced *S*s for their first response in the original alley, as compared with 12.6 required by *S*s run in the original alley only. The original-only *S*s required only 2.0 trials to make two additional goal responses; the total of 14.6 may be compared with the 15.0 trials just reported as required by the displaced *S*s for their first three goal responses.

At any stage in recovery, more displaced than original *S*s have operated the goal device. The displacement effect previously reported by Miller and Kraeling (**3**) and by Murray and Berkun (**5**) is confirmed here. Fewer displaced trials are required for recovery of the displaced response than original trials are for the original response. Error variance, however, has reduced the reliability of the difference between the 12.6 trials required by the original group and the 9.4 trials required by the displaced group ($P = .25$).

Certain corrections can be applied to properly bring out the difference in trials required between the original-only and the displaced groups. The small, insignificant difference between these two groups in the rate of recovery from the second shock (before the displacement variable is applied) can be taken into account by computing for each the mean difference between trials to recovery from the second shock and from the third shock. The difference between these two mean differences is the net effect of displacement after the third shock, and is reliable at $P = .02$. This directly confirms Miller and Kraeling's finding that fewer trials are required for recovery of approach in a displaced situation.

An alternative correction assumes that sniffing at the panel over the food tray is an indication of a significant amount of recovery. The *E* had noted that all entries into the final 4-in. segment of the alley were characterized by some sniffing at the goal box, whether or not the food mechanism was actually operated.

The original-only group required an average of 8.7 trials for the first goal sniff to occur, compared with 3.3 trials required by the displaced group, a difference reliable at the .007 level. This again confirms the displacement effect in terms of trials required to return to the goal area.

Although they began to sniff at the shield sooner than did *S*s in the original alley, *S*s in the displaced alley seemed to require more of these sniffs before

actually moving the shield (difference = 2.25 trials; $P = .13$). These additional trials, during which they only sniffed, reduced the mean difference between the groups in number of trials to first complete a goal response. It seems plausible to assume that the food-delivery mechanisms may not have been as different from each other as were the rest of the alleys. Miller and Kraeling had used lights of different intensities to emphasize the differences between food mechanisms. Such lights were not used in the present experiment, and this difference may account for the less reliable displacement effect reported here when the measure used is number of trials to complete the first goal response.

Speed.—Figure 2 shows that the speed of running down the alley and operating the food mechanism increases with successive trials on which *S* makes a goal response. That is, recovery is not complete when the first goal response occurs. The increment in speed from Response 3 to Response 5 is reliable at better than the .001 level.

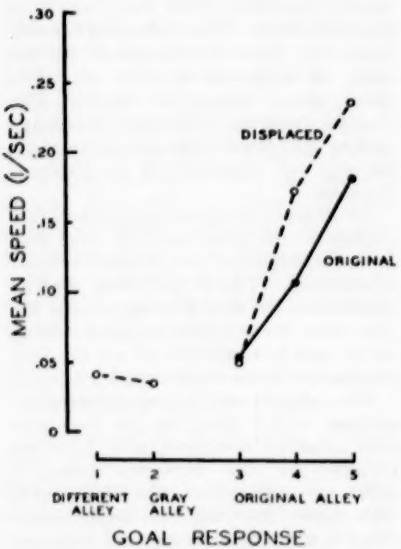

FIG. 2. Mean speed of approaching and operating food mechanism, for the five trials on which the displaced group responded and the three trials on which the original group responded.

When the displaced *S*s are shifted from one alley to another, their speed does not change appreciably. The therapeutic effect of a goal experience appears to be balanced by the loss in this fear reduction due to stimulus generalization decrement. But the effect of the displaced trials is not entirely lost. A persisting effect is indicated by the reliably greater speed ($P = .03$) of the displaced group on Responses 4 and 5 in the figure.

The Effect of the "Therapy"

On its first trial back in the original alley after responding in the intermediate alley, the displaced group traveled a mean distance of 12.2. This is reliably ($P < .001$) farther than the mean of 5.6 traveled by the original-only group on its first test trial.

The *S*s who have already made two displaced goal responses require only 2.6 trials to recover when tested in the original alley. This is reliably ($P < .001$) smaller than the 12.6 trials required for recovery by *S*s who were given no displaced experience.

But this clear effect of two goal responses in the different alley on the subsequent behavior in the original alley is smaller than the effect of the first two goal responses for the group that works in the original alley only. On the trial immediately following its first two goal responses in the original alley, the original-only group advances to a mean distance of 13.0. This is reliably ($P = .002$) farther than the mean distance of 12.2 reached by the displaced group on its first original-alley trial after its two

displaced goal responses. Similarly, all 32 of the original-only *Ss* make their third goal response immediately after their second, while only 9 of the 32 displaced *Ss* do this. This difference, representing a stimulus-generalization decrement in the shift from the different alley to the original alley, is reliable beyond the .001 level.

As a result, the total of 15.0 trials in both alleys required on the average by the displaced *Ss* for the occurrence of their first response in the original alley is not reliably different ($P = .4$) from the 12.6 trials required by the original-only *Ss* for their first response. Under the conditions of this experiment, the reduction in recovery time attributable to displacement therapy is approximately balanced by the stimulus-generalization decrement due to changing from the different back to the original alley. The displacement treatment is advantageous only if working time in the different situation is "cheaper" than working time in the original situation. This may also be a factor in keeping human clinical patients from leaving the therapy situation.

Future studies will require greater precision in measuring the relative slopes of the approach and avoidance generalization gradients. They should, therefore, use several degrees of difference between original and displaced situations, omit the intermediate stage of recovery, and measure the effect of less displaced experience than that involved in a complete displaced goal response.

Effect of Omitting Reinforcement of the First Two Goal Responses

The *Ss* who made two nonreinforced responses in the displaced alley were compared with *Ss* whose first two displaced responses were reinforced. Despite the fact that the odor of food in the tray was generally present for both conditions, *Ss* reinforced on their first postshock response (in both original and displaced groups) had performed this first response reliably faster ($P = .01$) than did *Ss* not reinforced for this response. Apparently some cue, possibly a stronger odor, indicated the presence of food. There is no other indication that these groups differ in strength of approach.

When subsequently tested in the original alley, with the food present for all *Ss* so that food cues could not account for any difference, *Ss* who had been reinforced for the displaced responses also recovered from the conflict more quickly than did those whose reinforcements had been omitted. Both for distance traveled on the first original trial and for number of trials in the original alley required for recovery, the difference between these two groups is reliable at well past the .001 level.

The *Ss* whose first two goal responses were made in the original alley are so nearly completely recovered that no effect of giving or omitting reinforcement for the first two responses can be shown. All 32 of these *Ss* earned the maximum score of 13 on the first trial following the first two responses.

Intermittent versus Regular Reinforcement

Half of the rats had been trained to approach under regular reinforcement, the rest being reinforced on only 50% of their training trials. As would be expected from previous studies summarized by Jenkins and Stanley (**2**), RR rats ran faster during the training series ($P = .001$).

After the second shock, however, the IR group is superior to the RR

TABLE 1

RECOVERY FROM THE SECOND SHOCK BEFORE ADMINISTRATION OF THE THIRD SHOCK

	All *Ss*	Groups		Diff. IR−RR *P*
		IR	RR	
Per cent responding at once	67	81	53	.02*
Mean distance on Recovery Trial 1	11.3	12.3	10.3	.04†
Mean trials required to receive 3rd shock	2.51	1.47	3.56	.02†

* χ^2 test.
† Mann-Whitney *U* test.

group, as seen in Table 1. That is, intermittent reinforcement of the approach habit increased the resistance to disruption by shock much as other studies have shown it increased resistance to extinction by omission of reinforcement.

A small difference between these groups in rate of recovery from the third shock is consistent with the difference after the second shock. That is, IR rats tend to travel farther per trial and to reach the goal sooner, but the difference now fails of statistical reliability for all possible analyses.

There was no indication of a significant interaction between the IR-RR variable and the omission of reinforcement for the first two postshock goal responses. This suggests that the enhancing effect of the reinforcement in the displaced alley is not due merely to the prevention of experimental extinction. Such an extinction effect should have been reduced as a result of the intermittent reinforcement. And since the approach habit is known to be at asymptote, the most likely explanation of the superiority of the reinforced rats lies in counterconditioning, or *reciprocal inhibition* as it is called by Wolpe (**6**). The food presumably facilitates the reduction of fear of the goal (thus facilitating recovery of approach) by eliciting responses involved in eating which are incompatible with fear and block it out. Hall (**1**) has also found that the presence of a stimulus eliciting a response incompatible with fear enhances the reduction of fear. Moltz (**4**) gave "therapy" consisting of confinement in a portion of a runway to food, which had also been the site of shock. Fear reduction occurred because (presumably) there were fractional approach responses elicited by cues in the compartment. He found no reliable effect of feeding one subgroup during its confinement, possibly because the presence of food changed the situation so much that this actually constituted a displaced therapy, with, consequently, a reduced effectiveness.

SUMMARY

1. An approach-avoidance conflict was induced by shocking rats while they operated a food device at a distinctive end of an alley. Successive nonshock trials were then given until they resumed operating the device, half the rats first operating it in a new alley differing in width and color before being tested in the original alley.

2. On the first postshock trial, *Ss* in the different alley approached nearer the goal than did *Ss* in the original alley. They resumed making the goal response in the different alley after fewer trials than required in the original alley. These results confirm those of previous experiments on displacement.

3. When *Ss* who had first recovered in the different alley were brought back to the original alley, they then ran farther on their first trial, required fewer trials to recover the goal response, and ran more quickly to the goal device after recovery than did *Ss* without the different-alley recovery experience.

4. Two displaced goal responses, however, were less effective than were two original-alley goal responses in facilitating subsequent recovery in the original alley. The total number of trials in different *and* original alleys required for recovery in the original alley is slightly but not reliably greater than the number of trials required by nondisplaced *Ss* for recovery in the original alley. These results show that it is not

always most efficient to use recovery in a displaced situation as a step to recovery of the original goal response.

5. The *Ss* who had been trained to approach with intermittent reinforcement recovered from conflict in fewer trials than those trained with regular reinforcement.

6. Omission of food reinforcement for the displaced responses delayed subsequent recovery in the original alley.

7. The time required for a trial to end in a goal response diminished with successive trials, unless the alley in which *S* was tested was changed.

References

1. Hall, J. C. Some conditions of anxiety extinction. *J. abnorm. soc. Psychol.*, 1955, **51**, 126–132.
2. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
3. Miller, N. E., & Kraeling, D. Displacement: greater generalization of approach than avoidance in a generalized approach-avoidance conflict. *J. exp. Psychol.*, 1952, **43**, 217–221.
4. Moltz, H. Resistance to extinction as a function of variations in stimuli associated with shock. *J. exp. Psychol.*, 1954, **47**, 418–424.
5. Murray, E. J., & Berkun, M. M. Displacement as a function of conflict. *J. abnorm. soc. Psychol.*, 1955, **51**, 47–56.
6. Wolpe, J. Experimental neurosis as learned behavior. *Brit. J. Psychol.*, 1952, **43**, 243–268.

(Received July 5, 1956)

Journal of Experimental Psychology
Vol. 54, No. 1, 1957

MEANING ESTABLISHED BY CLASSICAL CONDITIONING[1]

CAROLYN K. STAATS AND ARTHUR W. STAATS

Arizona State College at Tempe

In recent times a number of psychologists, such as Cofer and Foley (**2**), Mowrer (**6**), and Osgood (**8**), to mention a few, have come to view meaning as a response—an implicit response with cue functions which may mediate other responses. Osgood (**8**) does not consider the meaning response elicited by the sign of an object to be the same response which the object elicits. Only a fraction of the total response made to an object can come to be elicited by the sign. Osgood calls this component of the total response "detachable." It can be "attached" to another stimulus, the sign, through conditioning, whereas the other components of the total response can only be elicited by the object itself. A number of studies (**1, 10, 12**) lend themselves readily to the interpretation of meaning as an implicit, mediating response.

If meaning is to be considered a response, however, the same expectations should apply to meaning as to other responses. It would be expected, consequently, that meaning could be classically conditioned, i.e., the meaning response elicited by a word would be conditioned to any contiguously presented stimulus.

It follows from the above that if a nonsense syllable was presented and immediately followed by a meaningful word, it would be expected that the meaning response elicited by the word would be conditioned to the nonsense syllable. Experimentally this might be difficult to demonstrate since one conditioning trial might not establish a sufficiently strong association between the nonsense syllable and meaning response to be measurable. It would be expected that multiple paired presentations would be necessary in order to establish a conditioned meaning response strong enough to be measured. However, if the same word was paired many times with the nonsense syllable, the fact that the nonsense syllable would come to elicit the same meaning could be accounted for on the basis of a direct association between the nonsense syllable and the word itself. The direct association of syllable and word could be eliminated, however, even with multiple syllable-word pairings. This would be possible by pairing the syllable on each trial with a different word, all of the words having, however, an identical or similar meaning component. These words would not have to be synonyms, if the following rationale is correct. If the total meaning response to a word is composed of response components which are "detachable," i.e., could be separately conditioned, it would be possible to use words with largely different meaning, but having a common component. For example,

[1] The authors for some time have been mutually interested in language behavior. This article, the first study in a research project on language jointly planned by the authors, is based on the theoretical-experimental method developed by the second author. They are presently conducting further studies in this research project. The experimental data in the present study obtained by the first author forms a portion of the results in the dissertation submitted by her in partial fulfillment of the requirements for the Ph.D. degree at the University of California at Los Angeles in 1957. The first author wishes to thank Professor John P. Seward for critical suggestions made on the dissertation.

the words PRETTY, SWEET, and HEALTHY have in common a positive evaluative meaning, and yet are quite different in meaning otherwise.

Thus on each conditioning trial the nonsense syllable could be paired with a different word and yet the same component of the meaning response would be evoked and associated to the syllable. If the nonsense syllable was never paired with the same word, a stable association between the syllable and the word would not be made. The association would be between nonsense syllable and meaning response. This process is schematized in more detail in Fig. 1 in the discussion section and will be dealt with further then.

In this study nonsense syllables were used as conditioned stimuli. The unconditioned stimuli were different words with a certain similarity in meaning. The hypothesis was that this similar component of meaning would be conditioned to the nonsense syllable with which the words were paired.

In order to test this hypothesis, it was necessary to use a method of measuring meaning. Osgood (**7**) has developed an instrument called the semantic differential, and Osgood and Suci (**9**) report that a large portion of the total variance in *Ss*' judgment of meaning can be accounted for in terms of three factors of meaning—evaluative, potency, and activity. Words that were heavily loaded with these three factors provided the three meaning responses to be conditioned in the three subexperiments included in the present study. The semantic differential provided the technique for measuring the meaning which was to be conditioned to the nonsense syllables.

Method

Subjects

The *Ss* were 86 students in elementary psychology at Arizona State College. They participated in the study to fulfill a course requirement. For 32 *Ss* the UCS's were words with high loading on the evaluative factor (Exp. I); 24 *Ss* had activity words as UCS's (Exp. II); and 30 *Ss* were conditioned with words with a high loading on the potency factor (Exp. III).

Procedure

Experiment I.—The *Ss* were run in groups. There were two groups with one-half of the *Ss* in each group. Two types of stimuli were used: nonsense syllables which were presented visually by slide projection on a screen, and words which were presented orally by *E*, with *Ss* required to repeat the word aloud immediately after *E* had pronounced it.

The *Ss* were first seated in a room so they could all see the screen and not see each other's papers. They were told that they were to be *Ss* in an experiment concerned with studying two different types of learning—to see the effectiveness of each. One learning task was to concern nonsense syllables, and the other words.

The *Ss* wrote their names on several sheets of blank paper and the first task began. The nonsense syllables were VAF, XAD, VEC, YIM, and GAH. The syllables were presented in random order, with exposures of 5 sec. The intervals between exposures were less than 1 sec. The *Ss* were instructed to relax between syllables, and not to think of anything. Each syllable was presented four times. At the conclusion, *Ss* were instructed to write down all of the syllables they could recall.

The *Ss* then had 33 words presented to them which they were to learn. Each word was presented orally by *E* one time with approximately 2-sec. intervals between words. After the word was presented by *E*, *Ss* were instructed to immediately repeat the word aloud and then to continue to pronounce the word to themselves until the next word was given. The words were of no special type. Examples are: AT, BRIEF, UNDER, and BY. After each word was presented once, *Ss* were instructed to write down all of the words they could recall. Then they were presented with 12 pairs of words. One of each pair was one that had just been presented. Their task was to recognize which of the two it was and write it down. An example was "BRIEF or BRIEFCASE."

These two tasks were presented to train *Ss* in the procedure and to orient them properly for

TABLE 1

SYLLABLE (CS)-WORD (UCS) PAIRS FOR GROUP 1, EXP. I

YOF–beauty	XEH–worthless	XEH–sick	WUH–note
LAJ–with	XEH–sour	LAJ–ship	WUH–stick
XEH–thief	QUG–the	LAJ–room	YOF–success
LAJ–car	XEH–enemy	XEH–stupid	QUG–sock
YOF–win	QUG–box	LAJ–deck	QUG–six
WUH–pen	QUG–clay	LAJ–mop	LAJ–the
GIW–key	LAJ–this	GIW–glass	GIW–side
QUG–chair	XEH–cruel	WUH–into	LAJ–light
LAJ–paper	QUG–sand	XEH–failure	LAJ–three
LAJ–cord	XEH–dirty	GIW–shoe	QUG–saucer
YOF–gift	YOF–sacred	XEH–disgusting	YOF–money
XEH–bitter	YOF–friend	YOF–happy	GIW–quilt
GIW–book	LAJ–leaf	YOF–pretty	LAJ–it
LAJ–letter	XEH–evil	WUH–glove	GIW–truck
YOF–sweet	WUH–string	XEH–agony	LAJ–ground
LAJ–in	QUG–and	GIW–cart	WUH–water
YOF–honest	QUG–dot	QUG–wheel	GIW–garage
GIW–radio	WUH–line	WUH–on	XEH–poison
XEH–ugly	WUH–train	WUH–sofa	QUG–twelve
WUH–four	YOF–valuable	QUG–dresser	GIW–ink
GIW–cup	LAJ–table	WUH–trunk	GIW–store
XEH–sad	WUH–can	XEH–fear	QUG–number
WUH–five	GIW–word	WUH–those	GIW–hat
YOF–smart	GIW–pencil	XEH–insane	GIW–eleven
QUG–up	YOF–steak	QUG–fork	WUH–shirt
WUH–pot	QUG–clock	QUG–eight	YOF–vacation
YOF–rich	GIW–of	YOF–healthy	YOF–love

the next phase of the experiment where the hypothesis was tested.

The *Ss* were then told that the primary purpose of the experiment was to study "how both of these types of learning take place together—the effect that one has upon the other, and so on." Six new syllables were used: YOF, LAJ, XEH, WUH, GIW, and QUG. The syllables were presented in the same way. Approximately 1 sec. after the syllable appeared on the screen *E* pronounced a word aloud. The intervals between presentations of syllables were again less than 1 sec. The *Ss* were told they could learn the syllables by just looking at them, but that they should simultaneously concentrate on pronouncing the words aloud and to themselves since there would be many words, presented only once.

The nonsense syllables were presented in random order, though never more than twice in succession, so that no systematic associations were formed between them. Each nonsense syllable was presented 18 times, and each time it was paired with a different word, i.e., there were 18 conditioning trials. A nonsense syllable was never paired with a word more than once, so that stable associations were not formed between a nonsense syllable and any word. Thus, 108 different words were used. Two of the syllables were always paired with words which had high loadings on evaluative meaning. Most of the relevant meaningful words were taken from Osgood and Suci (9). When appropriate words with high loadings could not be found in this way, a thesaurus supplied them. The other four syllables were paired with words which had no systematic meaning.

Table 1 illustrates the method. It contains the syllable-word pairs presented to Group 1 in Exp. I. For Group 1, XEH was paired with different words which had a negative evaluative meaning, and YOF was paired with words with a positive evaluative meaning. For Group 2, XEH was paired with the positive meaning words, and YOF with the negative meaning words, word order remaining constant.

When the conditioning phase was completed, *Ss* were told that *E* first wished to find out how many syllables they remembered. At the same time, they were told, it would be necessary to find out how they felt about the syllables since that might have affected how the syllables were learned. Each *S* was given a small booklet in which there were six pages. On each page was printed one of the nonsense syllables and a semantic differential scale. The scale was the 7-point scale which Osgood and Suci describe (9), with the continuum from

pleasant to unpleasant. An example is as follows:

QUG
pleasant :—:—:—:—:—:—:—: unpleasant

In the booklet, QUG was on the first page and the other syllables on the following pages: XEH, LAJ, WUH, YOF, and GIW, in that order. The *Ss* were told how to mark the scale and to indicate at the bottom of the page whether or not the syllable was one that had been presented.

The *Ss* were then tested on the words. Finally they were asked to write down anything they had thought about the experiment, especially the purpose of it, and so on, or anything they had thought of during the experiment. It was explained that this might have affected the way they had learned the task.

Experiment II.—The procedure was exactly repeated for these *Ss* except that the words used to condition meaning to XEH and YOF had high loadings on the activity factor. The "active" words used are as follows: fast, ferocious, tense, energetic, hot, brisk, agitate, speed, eager, sharp, quick, haste, fidgets, excited, young, hustle, frisky, spry. The "passive" words used are as follows: slumber, cool, listless, drowsy, loafing, dull, lazy, calm, old, slow, relaxed, sleep, resting, peaceful, inert, sluggish, lag, lifeless. Since all other conditions were identical to Exp. I, it is not necessary to completely list the syllable-word pairs.

The *Ss* were again divided equally into a Group 1 and a Group 2. For Group 1, YOF was paired with passive meaning words and XEH with active meaning words. This was reversed for Group 2. The semantic differential booklet was also the same except the syllables were judged on an active-passive dimension.

Experiment III.—The procedure was again the same, except that words with high loadings on potency meaning were used. The words used are listed as follows with "strong" words first and "weak" words second: powerful, athletic, sturdy, masculine, robust, healthy, heavy, rugged, brave, active, hard, loud, deep, sharp, rich, wide, thick, large; crippled, feeble, soft, frail, narrow, poor, dull, thin, cowardly, feminine, lame, fragile, delicate, sick, quiet, passive, small, shallow. The syllables were later judged on a strong-weak dimension. Group 1 had YOF paired with strong words, XEH paired with weak words; Group 2 had this reversed.

Design

The data for the three experiments were treated in the same manner. Three variables were involved in the design: conditioned meaning (pleasant and unpleasant, active and passive, or strong and weak, depending upon the experiment); syllables (XEH and YOF); and Groups (1 and 2). The scores on the semantic differential given to each of the two conditioned syllables were analyzed in a 2×2 latin square as described by Lindquist (**5**, p. 278) for his Type II design.

Results

All *Ss* were questioned about the purpose of the experiments. Of the 86 *Ss*, 9 indicated awareness of a relationship between certain words and syllables. For these *Ss* it could be suggested that any meaning which the syllables had acquired was due to this awareness. For this reason, the data were analyzed without the scores of the "aware" *Ss*. In order to maintain a counterbalanced design when these *Ss* were excluded, it was necessary to randomly eliminate three additional *Ss* from the data. The resulting *Ns* were as follows: 30 in Exp. I, 20 in Exp. II, and 24 in Exp. III.

Table 2 presents the means and *SD*'s of the meaning scores for Exp. I, II, and III. The table itself is a representation of the 2×2 design

TABLE 2

MEANS AND *SD*'S OF CONDITIONED MEANING SCORES

Exp.	Group	Syllables			
		XEH		YOF	
		Mean	*SD*	Mean	*SD*
I	1	4.80	1.80	2.40	1.50
	2	3.13	1.46	4.73	1.77
II	1	4.90	1.70	3.30	2.33
	2	3.00	1.79	5.00	2.00
III	1	4.42	2.14	6.33	.94
	2	4.58	2.25	3.92	2.32

Note.—The pleasant pole scored 1, unpleasant 7; passive 1, active 7; weak 1, strong 7.

TABLE 3

ANALYSIS OF VARIANCE OF CONDITIONING DATA

Source	Exp. I			Exp. II			Exp. III		
	df	*MS*	*F*	*df*	*MS*	*F*	*df*	*MS*	*F*
Between *Ss*									
Groups	1	1.66	.59	1	.10	.15	1	15.19	3.82
Error	28	2.82		18	3.54		22	3.98	
Within									
Cond. Meaning	1	60.00	20.62***	1	32.40	6.39**	1	20.02	4.24*
Syllables	1	2.40	.82	1	.40	.08	1	4.68	.99
Residual	28	2.91		18	5.07		22	4.72	
Total	59			39			47		

* $P < .06$.
** $P < .01$.
*** $P < .001$.

for each experiment. The pleasant extreme of the evaluative scale was scored 1, the unpleasant 7; the passive extreme was scored 1, active 7; for potency, weak was 1 and strong 7.

The analysis of the data for the three experiments is presented in Table 3. The results of the analysis indicate that the hypothesized conditioning effect occurred. In Exp. I the *F* for the conditioned evaluative meaning variable was significant at better than the .001 level. None of the other variables were significant.

In Exp. II the *F* for conditioned activity meaning was significant at better than the .05 level. None of the other variables were significant. In Exp. III the *F* for conditioned potency meaning was significant at better than the .06 level. The *df* in this case was only 1 and 22. None of the other variables were significant.

DISCUSSION

It was possible to condition components of the total meaning responses of words to contiguously presented nonsense syllables. This conception is schematized in Fig. 1, and in so doing, the way the conditioning in this study was thought to have taken place is shown more specifically. The nonsense syllable YOF, in this example, is presented prior to the word PRETTY. PRETTY elicits a meaning response. This is schematized in the figure as two component responses; an evaluative response r_{PV} (in this example, the words have a positive value), and the other distinctive responses that characterize the meaning of the word, R_P. The pairing of YOF and PRETTY results in associations between YOF and r_{PV}, and YOF and R_P. In the following presentations of YOF and the words SWEET and HEALTHY the association between YOF and r_{PV} is further strengthened. This is not the case with associations R_P, R_S, and R_H, since they occur only once and are followed by other associations which are inhibitory. The direct associations indicated in the figure between the nonsense syllable and

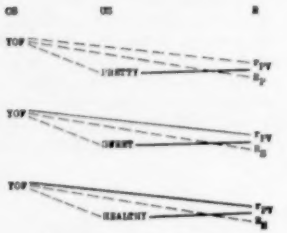

FIG. 1. The conditioning of positive evaluative meaning.

the individual words would also in this way be inhibited.

The results of this study have implications in several areas. (*a*) Using a conception of meaning as a mediating response, Mowrer (**6**) has suggested that a sentence is a conditioning device and that communication takes place when the meaning response which has been elicited by the predicate is conditioned to the subject of the sentence. The results of the present study substantiate Mowrer's approach by substantiating the basic theory that word meaning will indeed condition to contiguously presented stimuli. (*b*) Osgood (**8**) considers concept formation as learning a common mediating response for a group of stimuli. This study suggests that verbal concepts are signs that have been conditioned to the identical response components involved in the total response to each of several objects or signs. In Fig. 1, YOF is thus analogous to a concept; however, because of the design of the experiment, few *S*s attained this concept on a verbal level. (*c*) A question has arisen in the context of a number of studies of verbal behavior (**3, 4, 13, 14**), to name a few, concerning the necessity of awareness in order for learning to occur. The present results indicate that the meaning of stimuli may be learned without awareness. (*d*) Razran (**11**) has recently suggested that no American laboratory has produced second-order conditioning and that second-order conditioning "needs cognition to be adequately affected" (**11**, p. 329). However, in the present study, words, which are conditioned stimuli themselves, served as unconditioned stimuli in conditioning meaning to the nonsense syllables—and without cognition. (*e*) The results also suggest that there are psychological processes underlying the meaning factors arrived at by Osgood and Suci (**9**).

Summary

Three experiments were conducted to test the hypothesis that meaning responses elicited by a word can be conditioned to a contiguously presented neutral stimulus, e.g., a nonsense syllable. The study assumed that total word meaning is composed of response components which can be separately conditioned. A nonsense syllable was visually presented 18 times, each time paired with the auditory presentation of a different word. While these words were different, they all had an identical meaning component. In Exp. I, one nonsense syllable was paired with positive evaluative meaning and another was paired with negative evaluative meaning; in Exp. II "active" meaning and "passive" meaning responses were conditioned; and in Exp. III, "strong" and "weak" meaning responses were conditioned. In each experiment there was significant evidence that meaning responses had been conditioned to the nonsense syllables.

References

1. Bugelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.
2. Cofer, C. N., & Foley, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prologemena. *Psychol. Rev.*, 1942, **49**, 513–40.
3. Cohen, B. D., Kalish, H. I., Thurston, J. R., & Cohen, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.
4. Greenspoon, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
5. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
6. Mowrer, O. H. The psychologist looks at language. *Amer. Psychologist*, 1954, **9**, 660–694.
7. Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
8. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
9. Osgood, C. E., & Suci, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.
10. Razran, G. H. S. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–90.
11. Razran, G. H. S. A note on second-order conditioning—and secondary reinforcement. *Psychol. Rev.*, 1955, **62**, 327–332.

12. Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

13. Sidowski, J. B. Influence of awareness of reinforcement on verbal conditioning. *J. exp. Psychol.*, 1954, **48**, 355–360.

14. Thorndike, E. L., & Rock, R. T., Jr. Learning without awareness of what is being learned or intent to learn it. *J. exp. Psychol.*, 1934, **17**, 1–19.

(Received for early publication January 7, 1957)

ZEITSCHRIFT FÜR
EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Herausgeber für Österreich	Herausgeber für die Schweiz	Redaktion
PROF. DR. H. ROHRACHER, Wien	PROF. DR. R. MEILI, Bern	DR. C. J. HOGREFE, Göttingen

Band III/4

PETER R. HOFSTÄTTER, Dimensionen des mimischen Ausdrucks.

HANS CHRISTOPH MICKO, Untersuchungen der Tiefensehschärfe für größere Entfernungen.

J. RUTSCHMANN, Bemerkungen über mögliche Korrelationen zwischen visuellen Perzeptionen und elektrischen Phänomenen der Sehrinde.

HORST SACHER, Eintönige Lösungen bei Hebephrenen im Pyramidentest mit geordnetem Farbangebot.

HANS THOMAE, Untersuchungen über die Periodik im kindlichen Verhalten (Teil IV: Hans Thomae, Deutung der Befunde).

JOHANNES C. BRENGELMANN, Conrads Standardwahrscheinlichkeit in der Farbwahl: Eine Entgegnung.

HANS HÖRMANN, Zum Problem der psychischen Starrheit (Rigidität)

GEORG MEINECKE, Organisationstendenzen der wahrnehmungsnahen Reproduktion.

CARL-FRIEDRICH GRAUMANN, "Social Perception". Die Motivation der Wahrnehmung in neueren amerikanischen Untersuchungen (Sammelreferat).

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per number.
Subscription is DM 64.— per volume.

VERLAG FÜR PSYCHOLOGIE • DR. C. J. HOGREFE • GÖTTINGEN
(Brentanoweg 10, Germany)

AUTOMATIC EQUIPMENT FOR

BEHAVIORAL RESEARCH

FORINGER & CO.

AN IMPROVED
MULTIPLE STIMULUS UNIT

JUNE '57 CAT. #1166-4

A stimulus-generating and control unit which greatly simplifies the presentation of auditory and visual stimuli to the experimental subject; similar to our former No. 1166, but with 4 control relays, each associated with its own operating stud, manual switch, and pilot light. Three of the relays alter electrical values in an oscillator circuit; the fourth controls any 110 V. A.C. or 24 V. D.C. circuit. Output:

1. A tone of continuously adjustable pitch and volume.
2. A second tone of fixed pitch and adjustable volume.
3. Clicks of adjustable volume; rate adjustable from 2 per second to a rapid buzzing sound.
4. A ground (either 24 V. D.C. or 110 V. A.C.) to operate a lamp.

These four relays may be actuated singly or in any desired combination; moreover, selection and control may be done either manually, or remotely and automatically. As many as 15 different and useful stimuli may be obtained from this unit. All of them may be presented sequentially when operated in conjunction with our automatic programming apparatus. This Multiple Stimulus Unit is another of our standard clip-on building-block units.

Write for catalog sheet and price.

SEE OUR EXHIBIT AT THE
A.P.A. CONVENTION—BOOTH 35

FORINGER & CO. INC.

P.O. BOX 5882 BETHESDA 14, MD.